





Db 355 TCAPIPAGAEATFDLVNAQTG-----KVQTVTLTTDKNTVTYVGLDKNTYKFEVRS 407  
QY 318 AKPGSADLPENTNIATI-----NPNTSNDPQKQVTVRDGQITIKKIDGSTKASLOQAI 371  
Db 408 IKGSADYQEIITAGBIAVKNWKNDEPKPLDPTPEKAVTYGKKFKVKNDKNR--LAGAE 465  
QY 372 FVLKNA--TGQFL----- 382  
Db 466 FVIANADNAGQYLARKADKVKSOEEKQLVVTTKDALDRAVAAYNALTAQOQTQOKEKVDK 525  
QY 383 ---NFN-----DTNNVEWQTE---ANATEYTTGADGIIITGLKEGTYVLVEKKAPLGVN 431  
Db 526 AQAAYNAVAIANNAFADKONENVKLVSDAQGRFEITGLAGTYLLETKQAPAGYA 585  
QY 432 LLDSQKV-----ILGDGATDTTNSDNLVNPTEVNNKGTETLPSTGGIGTITFYIIG 483  
Db 586 LLTSRQKPEVATSYTSATQGIETAGSGK--DDATKVNKKITIPQTGGIGTIIFAVAG 643  
QY 484 ALLVIGAGIVLVA 496  
Db 644 AAIM---GIAVYA 653

## RESULT 7

US-10-282-122A-60689  
; Sequence 60689, Application US/10282122A  
; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 60689

; LENGTH: 793

; TYPE: PRT

; ORGANISM: Listeria monocytogenes

US-10-282-122A-60689

Query Match 9.0%; Score 232; DB 12; Length 793;

Best Local Similarity 25.1%; Pred. No. 1.1e-09;

Matches 125; Conservative 66; Mismatches 189; Indels 118; Gaps 23;

QY 45 TYKAYKVP--DAEDNANVSDNKGASLYLPQ-----GKEAEYK----- 82

Db 239 TYSADRVFTAGAPISSNFS--ATSDGFSVALGNLTDLSVQISYITTTTDDGKSTQYDNTAK 297

QY 83 -ASTDP-----NSLFTTTTNG-----GRYVYTKD--TASANEIATW-----A 117

Db 298 LAGTDFVTQTSTWTPASGGGGGGTGTGVTLTKEADAKTKATLEGAEFKLVDSKGTVLQ 357

QY 118 KSISANTPTSTVTESNNDGTEVINVSQYGYYSST-----VNNCAVIMVTSVTPNA 170

Db 358 ENITNAGSQLSIADLKFPDVTYQLIETKAPTGYKLDITTPVEFTIGENNOAI-----TVTKEN 413

QY 171 TIHEKNTDATWGDGKGK--TVDQKTVSVGDTVVKYITITYKNAVNVHGTGV-----YQYV 222

Db 414 TLNTGSVELTKLDAATKATLAGATFELQDKEGNTLQTDLTDENGVLKVTDLVPGSYQFV 473

QY 223 IKDTMPSASVVDLNEGSYEVITTDGSGNITTLTQSEKATGKYNLLENNNFTITIPWAA 282

Db 474 -ETSAPTGYKLDNSPVSEV--IAGETDQVVKVT-----KENTLEV-----SVELTK 518

QY 283 TNPPTGNTQNCANDDFYKGINITIVY-----TGVLK--SGAKPGSADLPEN---TNIATI 334

Db 519 LDSATKATLAGATFELQDKEGNTLQTDLTDENGVLKVTDLVPGSYQFVSAPTGYKLD 578

QY 335 NPNTS-----NDPQGVKTVRD---QGITIKKIDGSTKASLOGAIFVLKNATGQFLNF 384

Db 579 NSPVSEVAVAGETDQVVKVTENTLEVSGVSLTKDSATKATLAGATFELQDKEGNTLQ- 637

QY 385 NDTNNVWEGTEANATEYTTGADGIITITGLKEGTYVLVEKKAPLGNLLDNSQKVLGDG 444

Db 638 -----TGLTTDENGVLKVTDLVPGYQFVETKAPIGLYELDTTPVSPFEIVAG 683

QY 445 ATD---TTNSDNLVNPT 459

Db 684 ETDPVVKVTENTLVPPT 701

## RESULT 8

US-09-864-761-36047

; Sequence 36047, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aecmca-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30





```
; APPLICANT: Yang, Van-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-35

Query Match      8.5%; Score 219; DB 14; Length 915;
Best Local Similarity 22.5%; Pred. No. 1.5e-08;
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;

QY 29 AAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSKDGASYLIPQCKEAEYKASTD-- 86
Db 309 AKDNSNLTIQDSDAGNTDAKKV-----TFSNVSKDSKISASDHNVTLNSKVETSGDSTD 363
QY 87 -----FNSLFTTTTNGRGTYYTKKDTASAN-----EIAATWAKSISANTTP 126
Db 364 EDGNNNTGLTITAKNVTNNITSHKTVNITASENVTTKAGTTINATGTSVEVTAKTGD 423
QY 127 VSTVTESNNDGTEV-----INVSQYGYVYSSVTVNNGAVIMVTSVTPNATIHENKNTDAT 180
Db 424 IKGGIESGNNVNITASGDTLNVSNITQNVTVAAASGAVTTTKGSTINATTTGNANITTK 483
QY 181 WGDGGKTVDOQKTYSVGDTVK-YTITYKNAVNYHGTEKYQYVVKDTPMPSASVVDLNEGS 239
Db 484 TGEINGE-VKSAGNNVITASGNTLVNSNITQNVTVVANSAL--TTTEGSTINATTTGD 540
QY 240 YEVTTITDG-----SGNITTLTQSEKATGKYNLEENNNFTTIPWAAATNPTGNT 290
Db 541 ANITQTQNGINGKVESGSSVTLIATQTLAVGNIS-----GDTVTITADKGLTTQTSSK 596
QY 291 QNGANDDFFYKGI-----NTITVTYTVGL-----KSGAKPGSADLPENT 329
Db 597 INGT-----KSVTSSQSGDISGTSVTSATSLTTQAGSKIEAKTGEANVTSAT 650
QY 330 NI--ATINPNTSNDPPQKVTVRDQGITIKKIDGSTKASLOGAIFVLKNATGQFLNFNDT 387
Db 651 GTIGTISGNTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGTLTTETS 700
QY 388 NNVENG-----TEANATEYTTGADGIITITGLK-----EGTYVLVEKKA 426
Db 701 SDITSSNGQTLTAKDSSSIAGSINAANVTLLNTTGT--LTTVAGSKIEAASGTLVINAKDA 758
QY 427 PLGYNLLDNSQKVLGDGATDTTNSDNLNVNPTVENNKGTELPSTGGIGTTTFIIGAIL 486
Db 759 QL-----DGA---ASGDHTVVNATNANGSGSVIATT-----SSRVNITGDLI 797
QY 487 VIGAGIVLVAR 497
Db 798 TIN-GLNIISK 807

RESULT 11
US-10-193-764-37
; Sequence 37, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Van-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; NUMBER OF SEQ ID NOS: 91
```

```
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1222
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-37

Query Match      8.5%; Score 219; DB 14; Length 1222;
Best Local Similarity 22.5%; Pred. No. 2.3e-08;
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;

QY 29 AAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSKDGASYLIPQCKEAEYKASTD-- 86
Db 616 AKDNSNLTIQDSDAGNTDAKKV-----TFSNVSKDSKISASDHNVTLNSKVETSGDSTD 670
QY 87 -----FNSLFTTTTNGRGTYYTKKDTASAN-----EIAATWAKSISANTTP 126
Db 671 EDGNNNTGLTITAKNVTNNITSHKTVNITASENVTTKAGTTINATGTSVEVTAKTGD 730
QY 127 VSTVTESNNDGTEV-----INVSQYGYVYSSVTVNNGAVIMVTSVTPNATIHENKNTDAT 180
Db 731 IKGGIESGNNVNITASGDTLNVSNITQNVTVAAASGAVTTTKGSTINATTTGNANITTK 790
QY 181 WGDGGKTVDOQKTYSVGDTVK-YTITYKNAVNYHGTEKYQYVVKDTPMPSASVVDLNEGS 239
Db 791 TGEINGE-VKSAGNNVITASGNTLVNSNITQNVTVVANSAL--TTTEGSTINATTTGD 847
QY 240 YEVTTITDG-----SGNITTLTQSEKATGKYNLEENNNFTTIPWAAATNPTGNT 290
Db 848 ANITQTQNGINGKVESGSSVTLIATQTLAVGNIS-----GDTVTITADKGLTTQTSSK 903
QY 291 QNGANDDFFYKGI-----NTITVTYTVGL-----KSGAKPGSADLPENT 329
Db 904 INGT-----KSVTSSQSGDISGTSVTSATSLTTQAGSKIEAKTGEANVTSAT 957
QY 330 NI--ATINPNTSNDPPQKVTVRDQGITIKKIDGSTKASLOGAIFVLKNATGQFLNFNDT 387
Db 958 GTIGTISGNTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGTLTTETS 1007
QY 388 NNVENG-----TEANATEYTTGADGIITITGLK-----EGTYVLVEKKA 426
Db 1008 SDITSSNGQTLTAKDSSSIAGSINAANVTLLNTTGT--LTTVAGSKIEAASGTLVINAKDA 1065
QY 427 PLGYNLLDNSQKVLGDGATDTTNSDNLNVNPTVENNKGTELPSTGGIGTTTFIIGAIL 486
Db 1066 QL-----DGA---ASGDHTVVNATNANGSGSVIATT-----SSRVNITGDLI 1104
QY 487 VIGAGIVLVAR 497
Db 1105 TIN-GLNIISK 1114

RESULT 12
US-10-193-764-34
; Sequence 34, Application US/10193764
; Publication No. US20030133943A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Van-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
; FILE REFERENCE: 1038-1239MIS
; CURRENT APPLICATION NUMBER: US/10/193,764
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 09/167,568
; NUMBER OF SEQ ID NOS: 91
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-193-764-34

Query Match      8.5%; Score 219; DB 14; Length 1228;
Best Local Similarity 22.5%; Pred. No. 2.4e-08;
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;

QY 29 AAEFTGTTVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQGEAEYKASTD-- 86
Db 622 AKDNSLTIGDSDNAGTDAKKV-----TFSNVKDSKISASDHNVTLSKSVTSGBDST 676

QY 87 -----FNSLFTTTNGRVTYKDDTASAN-----BIATWAKSISANTTP 126
Db 677 EDGNNNTGLTITAKNVTVNNNITSHKTNITASENVTTKAGTTINATGSEVETAKTGD 736

QY 127 VSTVTSNNDGTEV-----INVSQYGYVSVSTVANGAVIMVTSVTPNATHEKNNTDAT 180
Db 737 IKGIESNSGNVNITAGSDTLNVSNTGQNVVAAASGAVTTTKGSTINATGNANITTK 796

QY 181 WDGGGKTVDOQKYSYGVDTVK-YTITYKNAVNYHGTKEYVQYVVKDTPSASVVDLNEGS 239
Db 797 TGEINGE-VKSASGNVNITASGNLTNVSNTGQNVTVTANSGLA---TTTEGSTINATTGD 853

QY 240 YEVITIDG-----SGNITLTGSEKATCKYNLLENNNFTITIPWAATNPTGNT 290
Db 854 ANITTQTNINGKVESSGSVTLIATGQTLAVGNIS-----GDTVTITADKGLTTQTSSK 909

QY 291 QNGANDDFYKGI-----NTITVYTGVL-----KSGAKPGSADLPENT 329
Db 910 INGT-----KSVTSSQSGDISGTTISGNTVSVATGSLTTQAGSKI EAKTGEANVTSAT 963

QY 330 NI--ATINPNTSNDPQKVTRDGOITIKIDGSTKASLQGAIFVLKNATGQFLNPNFT 387
Db 964 GTTGGTISGNTVN-----VTANTDNLTIK--DGAIRIKATGGAVTL--TATGGTLTTETS 1013

QY 388 NNVEWG-----TEANATEYTTGADGIITIGLK-----EGTVYVEKKA 426
Db 1014 SDITSSNGQTLTAKDSSTAGSINAANVTNLTGT--LTTVAGSKIEAASGLVLVINAQDA 1071

QY 427 PLGYNLLDSSQKVLGDGATDTTNSDNLVNPVTENNKGTELPSTGGIGTITFYIIGAIL 486
Db 1072 QL-----DGA---ASGDHTVNVNATNANGSGSVIATT-----SSRWNIITGDLI 1110

QY 487 VIGAGIVLVAR 497
Db 1111 TIN-GLNIISK 1120

RESULT 13
US-10-369-493-13955
; Sequence 13955, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 13955
; LENGTH: 1465
; TYPE: PRT
```

```
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13955

Query Match      8.5%; Score 218; DB 15; Length 1465;
Best Local Similarity 21.7%; Pred. No. 3.7e-08;
Matches 146; Conservative 61; Mismatches 207; Indels 260; Gaps 30;

QY 11 VASLAFGMAVSPVTPIAFAAETGTTT-VQDTQKGATYKAYKVFDAEI-----DNAN 60
Db 146 LVTFESEA VGTWNADLITANGTILSAVSSDGGITWTGTTFTPTASITDTTNLISLDNTG 205

QY 61 VSD-----SNKDGASYLIPQGEAEYKASTDFN-----SLFTTTNGRVTYVTKKT 107
Db 206 IADLAGNAGSGTTDSANVVIDTVRPTATVWADSNLAAGETSLVITITFSEAVSGFTADL 265

QY 108 ASANEI-----ATWAKSISANTTPSVSTVTSNN-----DGTEVI-----N 142
Db 266 TVANGTLSALSSDGGITWTVTL-----TPTSNTITDNLITLTDNTGVVDLACNAGSGTTN 321

QY 143 VSOY-----GYVVSSTVANGAVIMVTS-- 165
Db 322 SNNYAIDTARPTATVWADSNLRIGETSLVITTFSEAITGLTNADLTIANGLTAVSSD 381

QY 166 -----VTPNATHEKNNTDAT-----WGGGGKTVDOQKYS-- 195
Db 382 GGIWTATFTPTASI-----TDATNLITLNTGTIADLNGNAGSGTTDSNNYAITDVRPTAT 437

QY 196 -----VGDVVKYTTYKNAVNYHGTKEYVQYVVKDTM-----PS 229
Db 438 IVVANNLRIGETSLVITTFSEAVS--GFTNADLTVANGTLSALSSDGGITWTATFTPS 495

QY 230 ASVVDL-----NEGSYEVTITDGGG-----NIT 252
Db 496 ASVTDTNLITLTDNTGIADLAGNAGSGTTDSNNYAITDQRPATVWADSNLSAGETSIV 555

QY 253 TLTQSEKATCKYNLLENNNFTIT-----IPWAATNPTGNTQNGANDDFYKGIN 304
Db 556 TIT-FSEAVTGTNADLTVANGTLSAVSSDGGITWTATFTP-----NVGVND-----ATN 605

QY 305 TITVTYTVGLKSGAKPGSADLPENTNINATINPNTSND-----PGQKVTVRDGOITI-- 356
Db 606 LITLANTGI-----ADLSGNTSGTTSNNYSIDTLPATVIVADNALKIGETSL 656

QY 357 -----KKIDGSTKASLQGAIFVLKNATGQFLNFDNTNNVEWGTENATEYTTGADGIIT 410
Db 657 VTITFSEAVTGTNADL-----TIANGTLSAVSSD--GGITWTATFTPTTSTIDATNLIT 710

QY 411 ITGLKEGYTVLVEKAPLGYNLLDNS-QXVILGDGATDTTNSDNLVNPVTENNKGTELP 469
Db 711 -----LDNSGVQLNSGAGSGTTDSNNYAITD--TVRPTATMLVA 747

QY 470 STG-GIGTTTFYII 482
Db 748 DTALGIGQTLVTI 761

RESULT 14
US-10-282-122A-73634
; Sequence 73634, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
;; FILE REFERENCE: ELITRA.034A  
;; CURRENT APPLICATION NUMBER: US/10/282,122A  
;; CURRENT FILING DATE: 2003-02-20  
;; PRIOR FILING DATE: 2003-02-20  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR FILING DATE: 2000-03-21  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR FILING DATE: 2001-02-16  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: Patent in version 3.1  
;; SEQ ID NO 73634  
;; LENGTH: 645  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (569)..(569)  
;; OTHER INFORMATION: X=any amino acid  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (586)..(586)  
;; OTHER INFORMATION: X=any amino acid  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (617)..(617)  
;; OTHER INFORMATION: X=any amino acid  
;; FEATURE:  
;; NAME/KEY: MISC FEATURE  
;; LOCATION: (641)..(641)  
;; OTHER INFORMATION: X=any amino acid  
US-10-282-122A-73634

Query Match 8.2%; Score 212; DB 12; Length 645;  
Best Local Similarity 23.3%; Pred. No. 3.4e-08;  
Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;  
QY 34 TITVQDTQKATYKAYKVFDAIDN-----ANVSDSKDGASYLI 73  
DB 26 TTTVQDKGKNGKGRAPKIKVEDITSPSIRRDAAATPRNGIRVTVYDNDNGV-- 83  
QY 74 POCKEAYKASTDFNSL-----FTTTNGGRTYVTKDTASANEIATWAKSISANT 124  
DB 84 DGVQVLSNKDIYNGIDGDSAPITTTKDKGDGHTITVQNPDSSTTVVKGKQK 143  
QY 125 TPVSTVTESSNDGTEVINVSQYGYVSVTVANGAVIMVTSVTPNATIEKN-----TDA 179  
DB 144 TANITITTE-NPDGSHITITVNPDSGSKETVKNKG-----DGKTPKVEVTDNDNGHTVVKV 198  
QY 180 TWGDDG-----GKTVQKIVSGDVTVKYITIKYQVNVYHTEKVOYVVKD--- 225  
DB 199 TDGQGNVTNAILKDKGDKGAATATTENPDG-SHTVITNP---DGTKN--EFVVKNGRD 252  
QY 226 ----TPMSASVDLNGESYEVTITDGSNITLTQSEKATGKYNLLENNNNFTTIPWA 281  
DB 253 GVDGRPTASVRDNGDGSHTIITNPEG-VTTETTVRQKSPKVTITDE----- 300

QY 282 ATNPTGNTQNGANDDFYKGINITITVYTGVLKSGAKPGSADLPENTNIATINPTNSD 341  
DB 301 -----QNGTHKISVLNGDGTITET---IIKDKSPVATVRDNDQDGTITIRVENGNG 348  
QY 342 DPGQKVTVRDGO-----ITIKIDGSTRASLQGAIFVLKKNATGQFLNFND 386  
DB 349 TVSE-TTVRDGKSPAKVVDNGDGTHTITVNSDGIITTT-----TVRDGPEPKLEVID 401  
QY 387 TNNVEWGTEANATEYTTGADGIITITGLKEGTYLVKAPLGNLLDNSQ----- 437  
DB 402 NN-----DGSHTIKVTGADGKGTITITFDG-----KSPKA-NIVDNGDGTHTLTIVD 447  
QY 438 -----KVLGCGATDTTNSDNLVNPV---ENKKGTEL 468  
DB 448 SDGREYKSIKDG-KDGKDS-----VSPTVTVKNNDGTHV 482  
RESULT 15  
US-10-282-122A-74083  
; Sequence 74083, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; LENGTH: 2551  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-74083

Query Match 8.2%; Score 212; DB 12; Length 2551;  
Best Local Similarity 23.3%; Pred. No. 2.5e-07;  
Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;  
QY 34 TITVQDTQKATYKAYKVFDAIDN-----ANVSDSKDGASYLI 73

Db 1891 TPTVKGKNGKGRAPKIKVEDITSPSIRRRDAAAATPRNGIRVTVVDDVNDNGVY-- 1948  
QY 74 PQKEAEYKASTDFNSL-----FTTTNGGRVTVTKDTASANEIATWAKSISANT 124  
Db 1949 DEVDKVLNSKDIYNGIDGRDSAPITITKNDGDGHTITVQPDGSESTTVVKDGKQK 2008  
QY 125 TPVSTVTESSNDGTEVINVSQYGYVVSFVNNGAVIMVTSVTPNATIHKN-----TDA 179  
Db 2009 TANITITTE-NPDGSHITITVNPDPGSKETVVKNGK-----DGKTPKVEVTDNNDGTHVKV 2063  
QY 180 TWGDGG-----GKTVDKTVSVGTVKTYITTKNANVYHGTETKVYQVVKD--- 225  
Db 2064 TDGSGNVTNAILIKDGDGAATATTTTENPDG-SHTVTITNP-----DGTKN--EFVWKNGRD 2117  
QY 226 ----TWPSASVLDNLSGSEVITITDGSNITITTOGSEKATGYNLLENNNFTIIPWA 281  
Db 2118 GVDGRPTASVRNGDGSHTIVITNPEG-VTETTVRDGSKPKVITDE----- 2165  
QY 282 ATNPTGTONGANDFFYKGINITITVTVGLKSGAKPSADLPENTNIATINPNTSND 341  
Db 2166 -----QNGTHKISVLNGDGTETET---IIKDGKSPVATVRDNDQDGTYIRVENGG 2213  
QY 342 DPQCKTVTRDQ-----IITIKIDGSTKASLOGAIFVLKNATQOFLNPD 386  
Db 2214 TVSE-TTVRDGKSPTAKVDNDGSHITITVNSDGIITTT-----TVRDGREPKLEVID 2266  
QY 387 TNNVWGTEANATEYTTGADGIITITGLKEGTYVLVEKKAAPLGYNLLDNSQ----- 437  
Db 2267 NN-----DGSHTIKVTGADGKGTITITPDG-----KSPKA-NIVDNGDGHITITIVD 2312  
QY 438 -----KVILGDGATITTSNDLLNPTV---ENNKGTGL 468  
Db 2313 SDGREYKSIKDG-KDGKDS---VSPTVTVKNNDGTHV 2347

RESULT 16  
US-10-661-809-11  
; Sequence 11, Application US/10661809  
; Publication No. US20040101919A1  
; GENERAL INFORMATION:  
; APPLICANT: HOOK, Magnus  
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR  
; FILE OF INVENTION: POSITIVE BACTERIA  
; FILE REFERENCE: P07741US01/BAS  
; CURRENT APPLICATION NUMBER: US/10/661,809  
; PRIOR FILING DATE: 2003-09-15  
; PRIOR APPLICATION NUMBER: 60/410303  
; PRIOR FILING DATE: 2002-09-13  
; NUMBER OF SEQ ID NOS: 24  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 11  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-661-809-11  
Query Match 8.1%; Score 208.5; DB 16; Length 476;  
Best Local Similarity 26.5%; Pred. No. 4.2e-08;  
Matches 85; Conservative 49; Mismatches 124; Indels 63; Gaps 14;  
QY 187 KTV--DQKTVSGDVKYITTYKNANVYHGTETKVYQVVKDTPSASVLDNLSGSEVIT 244  
Db 192 KTVLDKQGFNGQEHINYQLTQIPANILG-----YQ-----EFRL 227  
QY 245 TDGSGNITITLTOGS--EKATGK-----YNLLENNNFTIIPWAATNPTGTONGANDD 297  
Db 228 SDKADYTLTLPLPSIEVKVAGKTVTGYTLTQKHGFTL-----DPSIKDLQNFAN-- 278  
QY 298 FFYKGINITITVTVGLKSGAKPSADLPENTNIATINPNTSNDPQCKVTVRQDGIITK 357  
Db 279 -----QTMTVSYQMRLEKTAEPDTAINNEQLV-----TDKHTLTKRATVRTGKSFV 326  
QY 358 KIDG-STKASLOGAIFVLKNATQOFLNFDNTNNVWNGTE-ANATEYTTGADGIITITGLK 415

Db 327 KVDSENAKITLPEAVFIVKNOAGEYLN--ETANGYRWQEKALAKKFTSNQAGEFSVKGLK 385  
QY 416 EGYTVLVKKKAPLGYNLLDNSQKVLGPGATDTTNSDNLVNVPTVENNKGTG--LPSTG 472  
Db 386 DGQYFLEELISAPKGYLLNQTEIPFTVQKNSVATNGQRTAPLH--VINKVKESGFLPKTN 443  
QY 473 GIGTTIFVFIIGAILVIGAGIV 493  
Db 444 E-ERSIWLTIAGLLIIGMWVI 463  
RESULT 17  
US-09-769-787-126  
; Sequence 126, Application US/09769787  
; Publication No. US20030091577A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129WO  
; CURRENT APPLICATION NUMBER: US/09/769,787  
; CURRENT FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 126  
; LENGTH: 886  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-769-787-126  
Query Match 7.8%; Score 200.5; DB 10; Length 886;  
Best Local Similarity 20.6%; Pred. No. 4.5e-07;  
Matches 107; Conservative 69; Mismatches 164; Indels 179; Gaps 24;  
QY 69 ASYLIPQGGKEAYKASTDFNSLFTTTNGGRVTVTKDTASANEIATWAKSI-----SAN 123  
Db 456 AAFPVKEPKYSEKKAAG--YAVIGDPINGGYLW-----NWRESILAYPFNSN 501  
QY 124 TTPVSTVTESSNDGTEVINVSQYGYVVSFVNNGAVIMVTSVTPNATIHKNATDWDG 183  
Db 502 TAKIT-----NHGDPTR-----WYXGNIAPDGVDVFTVGIGING---DPGTDEATAT 546  
QY 184 GGGKTVDDQKT---YSVGDIVK-----YTIITKNAVNYHGTETKVYQVVKDTPSPAS 231  
Db 547 SFMQSISSKPENYTNVTDTTKILEQLNRYFHTIVTEKKSIENT-----ITDPMGE 597  
QY 232 VVDLNEGS-----YEVTITDGS-----GNITITLTOGSEK---AT 262  
Db 598 LIDLQLTGDRFPDPADYTLTANDGSRLENGQAVGGPQNDGLLKNAKVLDTTEKRIVT 657  
QY 263 GYNLLENNNFTIIPWAATNPTGTONGANDDP-----FYKGINITITVTVGLKSGA 318  
Db 658 GLYLGTDKVKVTLTYNV-----RLNDEFVSNKFY----- 685  
QY 319 KPGSADLPENTNIATINP-----NTSNDPQCKVTVRD-----GQITIKK 358  
Db 686 -----DTNGRTLLHPKEVEQNTVRDPPFK--IRDVRKYPEITISKEKLGDIETFK 735  
QY 359 IDGSTKASIQGAIFVLKNATQOFLNFDNTNNVWNGTEANAT--EYTTGADGIITITGLKE 416  
Db 736 VNKNDKKPLRGAVFSLOK-----QHPDYPDIYGAIDQNGTYQNVRTGDKGLTFKNLSD 789  
QY 417 GTYVLVEKKAPLGYNLLDNSQ-----KVILGD-----GATDTTNSDNLVNP 459  
Db 790 GKRYLPENSEPAGYKPVQNKPIVAFQIVNNGEVRDVTISVPODIPAGYEFTNDKHYITNBP 849  
QY 460 VENNKGTELPSTGGTGTITFIYIIGAILVIGAGIVLVARR 498



```

Db      362  ----SUTKGAVETKVDIDCATLEGAVFKIVD-----MNGNDVR-----TDLTT 402
;
Qy      404  GAGGIITITGLKEGYTVLVEKAPLGYNLLDN--SOKVILGDGADTNTSDNLLVNPTVE 461
|||:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      403  DKDKGISVSLRPGDYQFIETKAPKHDLNQNPINFTEVKESQTATASVTATNSLTKGAVE 462
;
Qy      462  NNNKETLPSTGGTGTFIFYIGAILVI 488
|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      463  LMKVDDID----GTT---LEGAIFKI 481
;

RESULT 20
US-09-946-374-243
; Sequence 243, Application US/09946374
; Publication No. US20030073129A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleon
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PCL1
; CURRENT APPLICATION NUMBER: US/09/946,374
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099792

```

10 LMFGLLHLEAATNSNETSTANTGSSVSSGASTATNSGSSVTSGVSTATISGSSVTS 69  
69 ASYLIPQCKEABEYKA---STDFNSLFTTTTNGRTVTKDTASANEIATMAKSISANTT 125  
70 NGVSVITNSEPHTTSSGISTATNSSEFSTASS-GISIATNSSESTSSGASTATN-SEST 127  
126 PVSTVTESNNDGTEVINVSQYGYVSVTVNANGAVIMTVTPNATIIHEKNTDATWGDGG 185  
128 PSSGASTVTSNGSSVTSSG-----ASTATNSSESTVSSRASTATNSSESTLSS-GAST 179  
186 GKTVDQKTVSYGCDTVKYITTYKNAVNYHGTEKYVYVVKDTPMSASVVDLNEGSYEVTIT 245  
180 ATNSDSSTSSGAS-----TATNSSESS-----TTSSGASTATNSSESS---TVS 219  
246 DSGNIT-----TLTQSEKATGKYNLLENNFTITPWAATNPTGTQNGANDDPFY 300  
220 SRATATNSSESTSSGASTATNSERTTSNAGT-----ATNSSESTSSGA----- 267  
301 KGINTIVTYTGVKSGAKPGSADLPENTNIATINPNTSNDPQGVTVRQGITIKKID 360  
268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSESTSSGASTATNSD 319  
361 GSTKASLOGAIFVLKNATGQFLNFNDTNVNGTEANATEYTTGADGIITITGLKEGIYY 420  
320 SSTSS--GA-----GTATN-SESTVSSGISTVTNSSEST-- 352  
421 LVEKKAPLGYNLLDONSQKVLGDGATDTNSDNLNLPNTVENNKGTLPST-GGIGT 476  
353 -----PSSGANTATNSSESTSSGANTATNSSESTVSSGASTATNSSESTSSGVST 404

RESULT 21  
US-10-206-915-310  
; Sequence 310, Application US/10206915  
; Publication No. US20040029221A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE OF INVENTION: ACIDS ENCODING THE SAME  
; FILE REFERENCE: P3430R1C513  
; CURRENT APPLICATION NUMBER: US/10/206,915  
; PRIOR FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: 10/052586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28

Query Match 7.7%; Score 198; DB 10; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;  
14 LAFGM-----AVSPVTPIAAETGTTVQDTQKATYKAYVFDPAEIDNANVSDSNKDG 68

QY

; Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-206-915-310

Query Match 7.7%; Score 198; DB 12; Length 596;

Best Local Similarity 22.6%; Pred. No. 4e-07;

Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVPIAAFAETGTTIVQDTQKGATYKAYKVFDAEIDNANVDSNKG 68

Db 10 LMFGLLLHLEAATNSNETSANTGSSVSSGASTATNSGSSVTSSGVSTATISGSSVTS 69

QY 69 ASYLIPOQKEAEYKA---STDNSLFTTTNGGRYVVKDTASANEIATWAKSISANTT 125

Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS-GISATATNSSESTTSSGASTATN-SESST 127

QY 126 PVSTVTESNNDGTEVINVSQYGYVYVSVTVNNGAVIMVTSVTPNATIEHKNTDATWGDGG 185

Db 128 PSSGASTVNSGSSVTSSG-----ASTATNSSESTVSSRASTATNSSESTLSS-GAST 179

QY 186 GKTVDQKYSVGDYVKYITITYKNVNYHGTEKYQYVVKIDTMPASVVDLNEGSEYVIT 245

Db 180 ATNSDSSTSSGAS-----TATNSSESS-----TTSSGASTATNSSESS---TVS 219

QY 246 DSGNIT-----TLTQSEKATGKYNLLENNFTITIPWAATNPTGNTONGANDDPFY 300

Db 220 SRASTATNSSESTTSSGASTATNSSESTTSSGASTATNSSESTTSSGASTATNSSEST 267

QY 301 KGINTITVTYTGVLKSGAPGSADLPENTNIATINPNTSNDPQGVVTRDQGIITIKKID 360

Db 268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSESTTSSGASTATNSD 319

QY 361 GSTKASLOGAIFVLKNATQGFNFNDTNVWEGTEANATEYTTGADGIITITGLKEGTY 420

Db 320 SSTSS--GA-----GTATN-SESSTVSSGISTVITNSSEST-- 352

QY 421 LVEKKAPLGNLLDNSOKVILGCGATDTTNSDNLVNPVTENNKGTELPST-GGIGT 476

Db 353 -----PSSGANTATNSSESTTSSGANTATNSSESTVSSGASTATNSSESTTSSGVST 404

RESULT 23

US-10-199-670-310

; Sequence 310, Application US/10199670

; Publication No. US20040033560A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C401

; CURRENT APPLICATION NUMBER: US/10/199, 670

; CURRENT FILING DATE: 2002-07-19

; PRIOR APPLICATION NUMBER: 10/052586

; PRIOR FILING DATE: 2002-01-15

; PRIOR APPLICATION NUMBER: 60/059263

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/059266

; PRIOR FILING DATE: 1997-09-18

; PRIOR APPLICATION NUMBER: 60/062250

; PRIOR FILING DATE: 1997-10-17

; PRIOR APPLICATION NUMBER: 60/063120

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: 60/063121

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: 60/063486

; PRIOR FILING DATE: 1997-10-21

; PRIOR APPLICATION NUMBER: 60/063540

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063541

; PRIOR FILING DATE: 1997-10-28

; PRIOR APPLICATION NUMBER: 60/063544

; PRIOR FILING DATE: 1997-10-28

; Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 310

; LENGTH: 596

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-199-670-310

Query Match 7.7%; Score 198; DB 12; Length 596;

Best Local Similarity 22.6%; Pred. No. 4e-07;

Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVPIAAFAETGTTIVQDTQKGATYKAYKVFDAEIDNANVDSNKG 68

Db 10 LMFGLLLHLEAATNSNETSANTGSSVSSGASTATNSGSSVTSSGVSTATISGSSVTS 69

QY 69 ASYLIPOQKEAEYKA---STDNSLFTTTNGGRYVVKDTASANEIATWAKSISANTT 125

Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS-GISATATNSSESTTSSGASTATN-SESST 127

QY 126 PVSTVTESNNDGTEVINVSQYGYVYVSVTVNNGAVIMVTSVTPNATIEHKNTDATWGDGG 185

Db 128 PSSGASTVNSGSSVTSSG-----ASTATNSSESTVSSRASTATNSSESTLSS-GAST 179

QY 186 GKTVDQKYSVGDYVKYITITYKNVNYHGTEKYQYVVKIDTMPASVVDLNEGSEYVIT 245

Db 180 ATNSDSSTSSGAS-----TATNSSESS-----TTSSGASTATNSSESS---TVS 219

QY 246 DSGNIT-----TLTQSEKATGKYNLLENNFTITIPWAATNPTGNTONGANDDPFY 300

Db 220 SRASTATNSSESTTSSGASTATNSSESTTSSGASTATNSSESTTSSGASTATNSD 267

QY 301 KGINTITVTYTGVLKSGAPGSADLPENTNIATINPNTSNDPQGVVTRDQGIITIKKID 360

Db 268 ---STATNSDSSTVSSGASTAT-----NSESSTSSGASTATNSSESTTSSGASTATNSD 319

QY 361 GSTKASLOGAIFVLKNATQGFNFNDTNVWEGTEANATEYTTGADGIITITGLKEGTY 420

Db 320 SSTSS--GA-----GTATN-SESSTVSSGISTVITNSSEST-- 352

QY 421 LVEKKAPLGNLLDNSOKVILGCGATDTTNSDNLVNPVTENNKGTELPST-GGIGT 476

Db 353 -----PSSGANTATNSSESTTSSGANTATNSSESTVSSGASTATNSSESTTSSGVST 404



```
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C464
; CURRENT APPLICATION NUMBER: US/10/201,858
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-201-858-310

Query Match          7.7%; Score 198; DB 12; Length 596;
Best Local Similarity 22.6%; Pred. No. 4e-07;
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVTPPIAFAAETGTTIVQDTQKGATYKAYVDAEIDNANVSDSNKDG 68
Db 10 LMFGLLLHLEATNSNETSTANTGSSVSSGASTATNSGSSVTSSGVSATISGSSVTS 69
QY 69 ASYLIPQKAEYKA---STDNLSLFTTTNGGRYTVVKOTASANEIATWAKSISANTT 125
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS-GISIATNSESTTSSGASTATN-SESST 127
QY 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIEHKNTDATWGDGG 185
Db 128 PSSGASTVTNNGSSVTSSG-----ASTATNSESTTSSRASTATNSESTLSS-GAST 179
QY 186 GKTVDQKTVSGDVTVKYTYTYKNAVNYHGTEKYVYQVVKIDTMPASVVDLNEGSYEVTIT 245
Db 180 ATNDSSTSSGAS-----TATNSES-----TTSSGASTATNSES-----TVS 219
QY 246 DGSNGIT-----TLTQSEKATGKNLLENNFTITIPWAATNPTGTONGANDDPFY 300
Db 220 SRASTATNSESSTTSSGASTATNSESRTTSGAGT-----ATNSESSTSSGA----- 267
QY 301 KGINTITVTVYGVLSKAGPGSADLPENTNIATINPNTNDPQGVTVRDCQIITKKID 360
Db 268 ---STATNDSSTSSGASTAT-----NSESSTSSGASTATNSESSTTSSGASTATNSD 319
QY 361 GSTKASLQAGAI FVLKNATQFLNFNDTNVWEWTEANATEYTTGADGIITITGLKEGTVY 420
Db 320 SSTSS--GA-----GTATN-SESSTVSSGISTVTVNSESST--- 352
QY 421 LVEKAPLGYLLDNRQKVLGDGATDTTNSDNLLVNPTVENNKGTELPEST--GGIGT 476
Db 353 -----PSSGANTATNSESSTTSSGANTATNSESSTVSSGASTATNSESSTTSSGVST 404

RESULT 24
US-10-205-890-310

; Sequence 310, Application US/10205890
; Publication No. US20040048334A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C519
; CURRENT APPLICATION NUMBER: US/10/205,890
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-205-890-310

Query Match          7.7%; Score 198; DB 12; Length 596;
Best Local Similarity 22.6%; Pred. No. 4e-07;
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVTPPIAFAAETGTTIVQDTQKGATYKAYVDAEIDNANVSDSNKDG 68
Db 10 LMFGLLLHLEATNSNETSTANTGSSVSSGASTATNSGSSVTSSGVSATISGSSVTS 69
QY 69 ASYLIPQKAEYKA---STDNLSLFTTTNGGRYTVVKOTASANEIATWAKSISANTT 125
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS-GISIATNSESTTSSGASTATN-SESST 127
QY 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIEHKNTDATWGDGG 185
Db 128 PSSGASTVTNNGSSVTSSG-----ASTATNSESTTSSRASTATNSESTLSS-GAST 179
QY 186 GKTVDQKTVSGDVTVKYTYTYKNAVNYHGTEKYVYQVVKIDTMPASVVDLNEGSYEVTIT 245
Db 180 ATNDSSTSSGAS-----TATNSES-----TTSSGASTATNSES-----TVS 219
QY 246 DGSNGIT-----TLTQSEKATGKNLLENNFTITIPWAATNPTGTONGANDDPFY 300
Db 220 SRASTATNSESSTTSSGASTATNSESRTTSGAGT-----ATNSESSTSSGA----- 267
QY 301 KGINTITVTVYGVLSKAGPGSADLPENTNIATINPNTNDPQGVTVRDCQIITKKID 360
Db 268 ---STATNDSSTSSGASTAT-----NSESSTSSGASTATNSESSTTSSGASTATNSD 319
QY 361 GSTKASLQAGAI FVLKNATQFLNFNDTNVWEWTEANATEYTTGADGIITITGLKEGTVY 420
Db 320 SSTSS--GA-----GTATN-SESSTVSSGISTVTVNSESST--- 352
QY 421 LVEKAPLGYLLDNRQKVLGDGATDTTNSDNLLVNPTVENNKGTELPEST--GGIGT 476
Db 353 -----PSSGANTATNSESSTTSSGANTATNSESSTVSSGASTATNSESSTTSSGVST 404
```

Db 268 ---STATNSDSSTVSGASTAT-----NBSSTSTSGASTATNSSESTTSSGASTATNSD 319  
QY 361 GSTKASLOGAIFVLKATGQFLNFNDTNVVEWTEANATYTTGADGIITITGLKEGTY 420  
Db 320 SSTTSS--GA-----GTATN-SESSTVSGISTVTNSSEST-- 352  
QY 421 LVEKKAPLGYNLLDNSOKVILGDGATDTTNSDMLLVNPTVNNKGTLPST--GGIGT 476  
Db 353 -----PSSGANTATNSSESTTSSGANTATNSSESTVSSGASTATNSSESTTSSGCVST 404

RESULT 25  
US-10-208-024-310  
; Sequence 310, Application US/10208024  
; Publication No. US20040048335A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: F3430R1C538  
; CURRENT APPLICATION NUMBER: US/10/208, 024  
; CURRENT FILING DATE: 2002-07-29  
; PRIOR APPLICATION NUMBER: 10/052586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28  
; Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-208-024-310

Query Match 7.7%; Score 198; DB 12; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVTPIAAETGIIITVQDTQKGATYKAYKVFDAEIDNANVDSNKG 68  
Db 10 LMFGLLLHLEAAVNSNETSANTSGSVSSGASTATNSGSSVSTSGVSTATISGSSVTS 69  
QY 69 ASYLIPQGEAEYKA---STDFSLETTTNGRTYVTKDASANEIATWAKSISANTT 125  
Db 70 NGVSIYVNEFPHTTSSGISTATNSSESTASS-GISATNSSESTTSSGASTATN-SEST 127  
QY 126 PVSVTESNNDGTEVINVSQYGYVSVSVNNGAVIMVTPNATIHKNTDATWGDGG 185

Db 128 PSSGASTVNSGSSVTSSG-----ASTATNSSESTVSSRATSTATNSSESTLSS-GAST 179  
QY 186 GKTVDQKYVSGDTVKYIITYKNANVYHGTEKYQYVVKDTPMPSASVVDLNEGSYEVTIT 245  
Db 180 ATNSDSSTTSSGAS-----TATNSSEST-----TTSSGASTATNSSEST--TVS 219  
QY 246 DSGNIT-----TLTQSEKATGKYNLLENNNFITITPWAATNTPTGNTQNGANDDPFY 300  
Db 220 SRATATNSSESTTSSGASTATNSSESTTSSGAGT-----ATNSSESTTSSGA----- 267  
QY 301 KGINTIITVTTCVLKSGAKPGSADLPENTNIATINPNTSNDPQKQVTVRDCQIITIKKID 360  
Db 268 ---STATNSDSSTVSGASTAT-----NSESSTTSSGASTATNSSESTTSSGASTATNSD 319  
QY 361 GSTKASLOGAIFVLKATGQFLNFNDTNVVEWTEANATYTTGADGIITITGLKEGTY 420  
Db 320 SSTTSS--GA-----GTATN-SESSTVSGISTVTNSSEST-- 352  
QY 421 LVEKKAPLGYNLLDNSOKVILGDGATDTTNSDMLLVNPTVNNKGTLPST--GGIGT 476  
Db 353 -----PSSGANTATNSSESTTSSGANTATNSSESTVSSGASTATNSSESTTSSGCVST 404

RESULT 26  
US-10-201-853-310  
; Sequence 310, Application US/10201853  
; Publication No. US20040053358A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, James  
; APPLICANT: Smith, Victoria  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: F3430R1C465  
; CURRENT APPLICATION NUMBER: US/10/201, 853  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: 10/052586  
; PRIOR FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 60/059263  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/059266  
; PRIOR FILING DATE: 1997-09-18  
; PRIOR APPLICATION NUMBER: 60/062250  
; PRIOR FILING DATE: 1997-10-17  
; PRIOR APPLICATION NUMBER: 60/063120  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063121  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/063486  
; PRIOR FILING DATE: 1997-10-21  
; PRIOR APPLICATION NUMBER: 60/063540  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063541  
; PRIOR FILING DATE: 1997-10-28  
; PRIOR APPLICATION NUMBER: 60/063544  
; PRIOR FILING DATE: 1997-10-28  
; Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 612  
; SEQ ID NO 310  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-201-853-310



Db 180 ATNSDSSSTSSGAS-----TATNSSS-----TTSSGASTATNSSS---TVS 219  
QY 246 DSGNIT-----TLTGSEKATGKYNLLLENNFTTIPWAAATNTPTGNTONGANDFFY 300  
Db 220 SRATATNSSSTSSGASTATNSSERTTNGAGT-----ATNSSSTSSGA----- 267  
QY 301 KGINTITVTGVLKSGAKPGSADLPENTNIATINPNTSNDPQGVTVRDGQITIKKID 360  
Db 268 ---STATNSDSSSTVSSGASTAT-----NBSSTTSSGASTATNSSSTSSGASTATNSD 319  
QY 361 GSTKASLOGAIFVLKNATGQFLNFNDNNVWTEANATEYTTGADGIITITGLKEGY 420  
Db 320 SSTSS--GA-----GTATN-SESSSTVSSGISTVTNSSEST-- 352  
QY 421 LVEKKAPLGYNLLDNSOKVILGDGATDTTNSDMLLVNPTVENNKGTPLPST--GGIGT 476  
Db 353 -----PSSGANTATNSSSTSSGANTATNSSSTVSSGASTATNSSSTSSGVST 404

## RESULT 29

US-10-063-513-100  
; Sequence 100, Application US/10063513  
; Publication No. US20030018172A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063, 513  
; CURRENT FILING DATE: 2002-05-01  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-513-100

Query Match 7.7%; Score 198; DB 12; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVTPIAFAAETGTTITVDQTKGATYKAYKVFDAEIDNANVSDSNKDG 68  
Db 10 LMFGLLLHLEAATNSNETSANTGSSVSSGASTATNSSSVTSSTSSGVSTATISGSVTS 69  
QY 69 ASVLIPOGKEAEYKA---STDFNSLFTTTTNGRTVTVTKDPTASANEIATWAKSISANTT 125  
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS-GISATATNSSSTSSGASTATN-SESST 127  
QY 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIHEKNIDATWGDGG 185  
Db 128 PSSGASTVTSNGSSVTSSG-----ASTATNSSSTVSSRASTATNSSSTLSS--GAST 179  
QY 186 GKTVDOKTVSGDVTVKYITTYKNAVNYHGTKEKYQYVVIKDTMPASVVDLNEGSYEVTIT 245  
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS-GISATATNSSSTSSGASTATN-SESST 127  
QY 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIHEKNIDATWGDGG 185  
Db 128 PSSGASTVTSNGSSVTSSG-----ASTATNSSSTVSSRASTATNSSSTLSS--GAST 179  
QY 186 GKTVDOKTVSGDVTVKYITTYKNAVNYHGTKEKYQYVVIKDTMPASVVDLNEGSYEVTIT 245  
Db 180 ATNSDSSSTSSGAS-----TATNSSS-----TTSSGASTATNSSS---TVS 219  
QY 246 DSGNIT-----TLTGSEKATGKYNLLLENNFTTIPWAAATNTPTGNTONGANDFFY 300  
Db 220 SRATATNSSSTSSGASTATNSSERTTNGAGT-----ATNSSSTSSGA----- 267  
QY 301 KGINTITVTGVLKSGAKPGSADLPENTNIATINPNTSNDPQGVTVRDGQITIKKID 360  
Db 268 ---STATNSDSSSTVSSGASTAT-----NBSSTTSSGASTATNSSSTSSGASTATNSD 319

QY 361 GSTKASLOGAIFVLKNATGQFLNFNDNNVWTEANATEYTTGADGIITITGLKEGY 420  
Db 320 SSTSS--GA-----GTATN-SESSSTVSSGISTVTNSSEST-- 352  
QY 421 LVEKKAPLGYNLLDNSOKVILGDGATDTTNSDMLLVNPTVENNKGTPLPST--GGIGT 476  
Db 353 -----PSSGANTATNSSSTSSGANTATNSSSTVSSGASTATNSSSTSSGVST 404

## RESULT 30

US-10-063-515-100  
; Sequence 100, Application US/10063515  
; Publication No. US20030018173A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3230R1C1  
; CURRENT APPLICATION NUMBER: US/10/063, 515  
; CURRENT FILING DATE: 2002-05-01  
; Prior Application removed - See File Wrapper or Palm  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 100  
; LENGTH: 596  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-063-515-100

Query Match 7.7%; Score 198; DB 12; Length 596;  
Best Local Similarity 22.6%; Pred. No. 4e-07;  
Matches 108; Conservative 66; Mismatches 207; Indels 96; Gaps 18;

QY 14 LAFGM-----AVSPVTPIAFAAETGTTITVDQTKGATYKAYKVFDAEIDNANVSDSNKDG 68  
Db 10 LMFGLLLHLEAATNSNETSANTGSSVSSGASTATNSSSVTSSTSSGVSTATISGSVTS 69  
QY 69 ASVLIPOGKEAEYKA---STDFNSLFTTTTNGRTVTVTKDPTASANEIATWAKSISANTT 125  
Db 70 NGVSIVTNSEFHTTSSGISTATNSEFSTASS-GISATATNSSSTSSGASTATN-SESST 127  
QY 126 PVSTVTESNNDGTEVINVSQYGYVVSSTVNGAVIMVTSVTPNATIHEKNIDATWGDGG 185  
Db 128 PSSGASTVTSNGSSVTSSG-----ASTATNSSSTVSSRASTATNSSSTLSS--GAST 179  
QY 186 GKTVDOKTVSGDVTVKYITTYKNAVNYHGTKEKYQYVVIKDTMPASVVDLNEGSYEVTIT 245  
Db 180 ATNSDSSSTSSGAS-----TATNSSS-----TTSSGASTATNSSS---TVS 219  
QY 246 DSGNIT-----TLTGSEKATGKYNLLLENNFTTIPWAAATNTPTGNTONGANDFFY 300  
Db 220 SRATATNSSSTSSGASTATNSSERTTNGAGT-----ATNSSSTSSGA----- 267  
QY 301 KGINTITVTGVLKSGAKPGSADLPENTNIATINPNTSNDPQGVTVRDGQITIKKID 360  
Db 268 ---STATNSDSSSTVSSGASTAT-----NBSSTTSSGASTATNSSSTSSGASTATNSD 319  
QY 361 GSTKASLOGAIFVLKNATGQFLNFNDNNVWTEANATEYTTGADGIITITGLKEGY 420  
Db 320 SSTSS--GA-----GTATN-SESSSTVSSGISTVTNSSEST-- 352  
QY 421 LVEKKAPLGYNLLDNSOKVILGDGATDTTNSDMLLVNPTVENNKGTPLPST--GGIGT 476  
Db 353 -----PSSGANTATNSSSTSSGANTATNSSSTVSSGASTATNSSSTSSGVST 404

Search completed: July 22, 2004, 10:34:40  
Job time : 49 secs

---

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 22:18:23 ; Search time 6024 Seconds  
(without alignments)  
10857.344 Million cell updates/sec

Title: US-10-009-254-1  
Perfect score: 1509  
Sequence: 1 atgaaaagaagaattatca.....gtcgtcgtttacgtttcttaa 1509

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sv.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1509	100.0	1509	1	AF485279	AF485279 Streptoco
2	1509	100.0	1509	6	AX476887	AX476887 Sequence
3	92.8	6.1	216959	10	AC116998	AC116998 Mus muscu
4	89.6	5.9	171104	2	BX649545	BX649545 Danio rer
5	87.2	5.8	170221	10	AC127421	AC127421 Mus muscu
6	86.4	5.7	213544	2	BX510939	BX510939 Danio rer
c 7	86	5.7	116807	2	BX890614	BX890614 Danio rer
c 8	85.2	5.6	146275	2	AL935272	AL935272 Danio rer
c 9	84	5.6	198743	2	BX530070	BX530070 Danio rer
c 10	83.4	5.5	155204	2	AC007926	AC007926 Trypanoso
c 11	82.4	5.5	94534	5	AL929250	AL929250 Zebrafish
c 12	80.6	5.3	110000	2	PFMAL13_07	Continuation (8 of
c 13	80.4	5.3	225581	2	BX537105	BX537105 Danio rer
c 14	80	5.3	10115	5	AF397467	AF397467 Ictalurus
c 15	79.8	5.3	308092	3	AY333070	AY333070 Drosophil
c 16	78.6	5.2	224391	2	AC137124	AC137124 Mus muscu
c 17	78	5.2	164399	3	PFMAL3P6	Z98551 Plasmodium
c 18	77.8	5.2	154160	2	BX571709	BX571709 Danio rer
c 19	77.8	5.2	177595	2	AL953867	AL953867 Danio rer
c 20	77.4	5.1	200634	5	BX004990	AL935062 Zebrafish
c 21	77.2	5.1	176010	5	AL935062	AL935062 Zebrafish
c 22	75.6	5.0	151341	5	AL929536	AL929536 Zebrafish
c 23	75.4	5.0	22398	5	FRU21723	AJ271723 Fugu rubr
c 24	75.2	5.0	245292	2	BX088589	BX088589 Danio rer
c 25	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 26	75	5.0	183357	2	BX569779	BX569779 Danio rer
c 27	74	4.9	199698	2	BX004991	BX004991 Danio rer
c 28	73.8	4.9	250029	3	AF014820	AF014820 Plasmodiu
c 29	73	4.8	666	8	AF413050S2	AF413051 Zee maye
c 30	73	4.8	199347	2	BX321884	BX321884 Danio rer
c 31	72.8	4.8	146907	2	AC141031	AC141031 Rattus no
c 32	72.6	4.8	173134	5	AL935282	AL935282 Zebrafish
c 33	72.4	4.8	157033	2	BX323881	BX323881 Danio rer
c 34	72.4	4.8	179553	2	AC024253	AC024253 Homo sapi
c 35	72.2	4.8	153751	3	AC116551	AC116551 Dictyoste
c 36	72.2	4.8	180668	2	AC020857	AC020857 Mus muscu
c 37	70.6	4.7	146415	5	AL929469	AL929469 Zebrafish
c 38	70.6	4.7	212722	2	AC098909	AC098909 Rattus no
c 39	70.2	4.7	2167	3	PRSTARPA	Z30339 P. reichenow
c 40	69.6	4.6	202182	2	AC120639	AC120639 Rattus no
c 41	69.6	4.6	274060	2	AC095184	AC095184 Rattus no
c 42	69.4	4.6	203271	2	BX322566	BX322566 Danio rer
c 43	69.2	4.6	11534	3	PFARP1PR	Y08926 P. falciparu
c 44	69	4.6	178664	2	BX511115	BX511115 Danio rer
c 45	69	4.6	180023	2	BX897685	BX897685 Danio rer
c 46	68.8	4.6	333321	2	AC116986	AC116986 Dictyoste
c 47	68.6	4.5	209387	2	AC096448	AC096448 Rattus no
c 48	68.2	4.5	14626	3	BMOFLCEG	M76430 Bombyx mori
c 49	68	4.5	253151	3	AE014842	AE014842 Plasmodiu
c 50	68	4.5	293650	1	AP004172	AP004172 Mycoplasma
c 51	67.8	4.5	146570	3	AC117072	AC117072 Dictyoste
c 52	67.8	4.5	234117	2	AC130985	AC130985 Rattus no
c 53	67.6	4.5	169049	2	BX322580	BX322580 Danio rer
c 54	67.2	4.5	206779	2	BX511037	BX511037 Danio rer
c 55	66.8	4.4	115758	9	AC104634	AC104634 Homo sapi
c 56	66.8	4.4	165118	2	AC119053	AC119053 Danio rer
c 57	66.2	4.4	125623	3	AC115599	AC115599 Dictyoste
c 58	66	4.4	187846	2	AC146995	AC146995 Arabacia p
c 59	65.8	4.4	185107	2	AC118513	AC118513 Rattus no
c 60	65.8	4.4	243570	2	AC125757	AC125757 Rattus no
c 61	65.6	4.3	110000	2	PFMAL13_24	Continuation (25 o
c 62	65.4	4.3	186935	2	AC022322	AC022322 Homo sapi
c 63	65.2	4.3	114226	2	AC004710	AC004710 Plasmodiu
c 64	65.2	4.3	196490	2	AC005507	AC005507 Plasmodiu
c 65	65.2	4.3	210950	10	AC129317	AC129317 Mus muscu





```
QY 901 AAGGGAATAAATACAAATCACAGTCACTTATACAGGAGTATTAAAGAGTGGAGCTAAACCA 960
Db 901 AAGGGAATAAATACAAATCACAGTCACTTATACAGGAGTATTAAAGAGTGGAGCTAAACCA 960
QY 961 GGTTCAGCTGATTTACAGAGAAATCAAAATTCAGGAGTGGAGCTAAACCA 1020
Db 961 GGTTCAGCTGATTTACAGAGAAATCAAAATTCAGGAGTGGAGCTAAACCA 1020
QY 1021 GATGACCCAGGTCAAAAAGTAAACAGTGGAGGATGGTCAAAATTTACTATATAAAAAAATTGAT 1080
Db 1021 GATGACCCAGGTCAAAAAGTAAACAGTGGAGGATGGTCAAAATTTACTATATAAAAAAATTGAT 1080
QY 1081 GGTTCACAAAAGCTTCATTACAAAGTGTCTATATTTGTTTAAAGAAATGCTACGGGTCAA 1140
Db 1081 GGTTCACAAAAGCTTCATTACAAAGTGTCTATATTTGTTTAAAGAAATGCTACGGGTCAA 1140
QY 1141 TTTCTAAATCTTTAAAGATACAAATTAACGTTGAATGGGCGACAGAGCTAATGCAACAGAA 1200
Db 1141 TTTCTAAATCTTTAAAGATACAAATTAACGTTGAATGGGCGACAGAGCTAATGCAACAGAA 1200
QY 1201 TATACACAGGAGCAGATGGTATAATTAACATTAACAGGCTTGAAAGAGGTACATACATAT 1260
Db 1201 TATACACAGGAGCAGATGGTATAATTAACATTAACAGGCTTGAAAGAGGTACATACATAT 1260
QY 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAAATCTCAGAGGTTATT 1320
Db 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAAATCTCAGAGGTTATT 1320
QY 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTAGTTAAACCAACTGTT 1380
Db 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTAGTTAAACCAACTGTT 1380
QY 1381 GAAAATAACAAAGGCTACTGAGTTGCTTCAACAGGTTGATTTGGTACAAACAATTTTCTAC 1440
Db 1381 GAAAATAACAAAGGCTACTGAGTTGCTTCAACAGGTTGATTTGGTACAAACAATTTTCTAC 1440
QY 1441 ATTATAGTGCATTTTATAGTAATAGGAGCAGATATCGTGTGTTGCTCGTCTGTTTAA 1500
Db 1441 ATTATAGTGCATTTTATAGTAATAGGAGCAGATATCGTGTGTTGCTCGTCTGTTTAA 1500
QY 1501 CGTCTCTTAA 1509
Db 1501 CGTCTCTTAA 1509

RESULT 2
AX476887 LOCUS AX476887 1509 bp DNA linear PAT 12-AUG-2002
DEFINITION Sequence 15 from Patent WO0212294.
ACCESSION AX476887
VERSION AX476887.1 GI:22216151
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
REFERENCE 1
AUTHORS Adderson,E. and Bohnsack,J.
TITLE Group b streptococcus polypeptides nucleic acids and therapeutic
compositions and vaccines thereof
JOURNAL Patent: WO 0212294-A 15 14-FEB-2002;
ST. JUDE CHILDREN'S RESEARCH HOSPITAL (US) ; University of Utah
Research Foundation (US)
FEATURES
source Location/Qualifiers
1..1509
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
1..1509
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
CDS
```

```
/protein_id="CAD43673.1"
/db_xref="GI:22216152"
/translation="MKKKMIOSLLVSLAFQMAVSPVPIAFAPAEETGTTITVODTQKGA
TYKAYKVFDAIDNANVSDSNKDGASYLIPQKEAEYKASTDFNSLFTTTTNGGRTYV
TKKDSANEIATWAKSISANTTPTVSTVESNNDGTENVINVOYGYVYSSVANGAV
IMVTSVTPNATIIHEKNTDATTGSGGKVDKQTSVGGTIVKTIITIKYNAVHTEKRV
YQYVKTDPMSASVVDLNBGSYEVTITDGSNITTLTQSGEKATGKYNLLBENNNTTI
TIPMAATNTPTGNTQNGANDFFYKINTITVYTVLTKSGAKGCSADLPENTNIATI
NPNTSDPQKQVTRDQKITIKKIDGSTRASLQCAIFVLKNFNDQFLNFDNTNNVEW
GTEANATYTTGADGIIITIGLKEGTIVLVEKAPLGVNLLDNQKIVLGGADTDTTN
SDNLLVNPTVENNKGTELPSGGIGTTFYIIGAILVIGAGILVARRRLRS"
```

## ORIGIN

```
Query Match 100.0%; Score 1509; DB 6; Length 1509;
Best Local Similarity 100.0%; Pred. No. 3e-238;
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAAAGAAAATGATTCATTCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
Db 1 ATGAAAAGAAAATGATTCATTCGCTGTAGTGGCGAGTTTAGCATTTGGTATGGCTGTA 60
QY 61 TCACCAAGTTAGCCGATAGCTTTTGGCGCTCAGACAGGACAAATTTACAGTTCAAGATACT 120
Db 61 TCACCAAGTTAGCCGATAGCTTTTGGCGCTCAGACAGGACAAATTTACAGTTCAAGATACT 120
QY 121 CAAAAGGCGCAACTATAAAGCATATAAAGTTTTTGTATGACAGAAATAGATAATGCAAA 180
Db 121 CAAAAGGCGCAACTATAAAGCATATAAAGTTTTTGTATGACAGAAATAGATAATGCAAA 180
QY 181 GTATCTGATTCGAATAAGATGGAGCTTCTTATTTAATTCCTCAAGGTAAGAGAGCTGAG 240
Db 181 GTATCTGATTCGAATAAGATGGAGCTTCTTATTTAATTCCTCAAGGTAAGAGAGCTGAG 240
QY 241 TATAAAGCTTCAACTGATTTTAAATTTCTTTTACGACCACTACTTAATGAGGAGGAGAACA 300
Db 241 TATAAAGCTTCAACTGATTTTAAATTTCTTTTACGACCACTACTTAATGAGGAGGAGAACA 300
QY 301 TATGTAACATAAAAAGATACATGCGTCAGCAATGAGATGCGACATGGGCTAAATCTATA 360
Db 301 TATGTAACATAAAAAGATACATGCGTCAGCAATGAGATGCGACATGGGCTAAATCTATA 360
QY 361 TCAGCTAATACATACACAGTTTCCACTGTTACTGAGTCAAAATATGATGGTACTGAGTT 420
Db 361 TCAGCTAATACATACACAGTTTCCACTGTTACTGAGTCAAAATATGATGGTACTGAGTT 420
QY 421 ATTAATGTTTCCCAATATGATATTTATTTCTGCTAGCACTGTTTAAATGAGAGCTGTA 480
Db 421 ATTAATGTTTCCCAATATGATATTTATTTCTGCTAGCACTGTTTAAATGAGAGCTGTA 480
QY 481 ATTATGTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAGAAATGATGTCGACA 540
Db 481 ATTATGTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAGAAATGATGTCGACA 540
QY 541 TGGGAGATGTTGTTGAAAACCTGTAGATCAAAAACGTTACTCGGTTGGTATACATGTC 600
Db 541 TGGGAGATGTTGTTGAAAACCTGTAGATCAAAAACGTTACTCGGTTGGTATACATGTC 600
QY 601 AAATATATCTATTACTTATAAGAAATGCAAGTCAATTTATGTCAGACAAAGAGTGTATCAA 660
Db 601 AAATATATCTATTACTTATAAGAAATGCAAGTCAATTTATGTCAGACAAAGAGTGTATCAA 660
QY 661 TATGTTTAAAGAGTACTATGTCATCTGCTTCTGCTAGTTGATTTGCAAGAGGCTTTAT 720
Db 661 TATGTTTAAAGAGTACTATGTCATCTGCTTCTGCTAGTTGATTTGCAAGAGGCTTTAT 720
QY 721 GAACTAATCTATTACTGATGGATCAGGGAATATTAACAATCTCAAGGTTTCGAAAAA 780
Db 721 GAACTAATCTATTACTGATGGATCAGGGAATATTAACAATCTCAAGGTTTCGAAAAA 780
QY 781 GCAACTGGAGTATAACCTGTTAGAGGAAATATAATTTTCCAGTACTATTTCCGTTGG 840
Db 781 GCAACTGGAGTATAACCTGTTAGAGGAAATATAATTTTCCAGTACTATTTCCGTTGG 840
```

```
QY 841 GCAGCTACCAATCTCCAAACCGGAATACTCAAAATGGAGCTAAATGATGACTTTTTTAT 900
Db 841 GCAGCTACCAATCTCCAAACCGGAATACTCAAAATGGAGCTAAATGATGACTTTTTTAT 900
QY 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGAGGAGTATTAAGAGTGGAGCTAAACCA 960
Db 901 AAGGGAATAAATAACAATCAAGTCACTTATACAGAGGAGTATTAAGAGTGGAGCTAAACCA 960
QY 961 GGTTCAGCTGATTTACAGAGAAATACAACATTCGACCATCAACCCCAATAGCAAT 1020
Db 961 GGTTCAGCTGATTTACAGAGAAATACAACATTCGACCATCAACCCCAATAGCAAT 1020
QY 1021 GATGACCCAGGTCACAAAGTAAACAGTGGAGTGGTCAAAATCTATAAAAAAATTGAT 1080
Db 1021 GATGACCCAGGTCACAAAGTAAACAGTGGAGTGGTCAAAATCTATAAAAAAATTGAT 1080
QY 1081 GGTTCACAAAAGCTTCATTACAGAGTGTCTATATTTGTTTTAAAGAACTCTACGGGTCAA 1140
Db 1081 GGTTCACAAAAGCTTCATTACAGAGTGTCTATATTTGTTTTAAAGAACTCTACGGGTCAA 1140
QY 1141 TTTCTAAACTTTAACGATACAATAACGTTGAATGGGCGACAGAGCTAATGCAACAGAA 1200
Db 1141 TTTCTAAACTTTAACGATACAATAACGTTGAATGGGCGACAGAGCTAATGCAACAGAA 1200
QY 1201 TATACACAGGAGCAGATGGTATAATTACCATACAGGCTTGAAGAAGGTACATACAT 1260
Db 1201 TATACACAGGAGCAGATGGTATAATTACCATACAGGCTTGAAGAAGGTACATACAT 1260
QY 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAACTCTCAGAAGGTTAT 1320
Db 1261 CTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAACTCTCAGAAGGTTAT 1320
QY 1321 TTAGGAGTGGAGCCACTGATACGACTAATTCAGATAAAGCTTTAGTTAAACCACTGTT 1380
Db 1321 TTAGGAGTGGAGCCACTGATACGACTAATTCAGATAAAGCTTTAGTTAAACCACTGTT 1380
QY 1381 GAAATAACAAAGGTAAGTGGTTCCTTCAACAGGTTGGTATGGTACAACAATTTCTAC 1440
Db 1381 GAAATAACAAAGGTAAGTGGTTCCTTCAACAGGTTGGTATGGTACAACAATTTCTAC 1440
QY 1441 ATTATAGTGCATTTTAGTAAATAGAGCAGGATCGTGTGTTGCTCGTCTGCTGTTA 1500
Db 1441 ATTATAGTGCATTTTAGTAAATAGAGCAGGATCGTGTGTTGCTCGTCTGCTGTTA 1500
QY 1501 CGTTCCTAA 1509
Db 1501 CGTTCCTAA 1509

RESULT 3
AC116998
LOCUS AC116998 216959 bp DNA linear ROD 08-NOV-2003
DEFINITION Mus musculus BAC clone RP23-16F6 from 18, complete sequence.
ACCESSION AC116998
VERSION AC116998.3 GI:24080765
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 216959)
Isak,A., Kozlowski,A., Spalding,L. and Mangiapanello,L.
The sequence of Mus musculus BAC clone RP23-16F6
Unpublished (2001)
2 (bases 1 to 216959)
Wilson,R.
Sequencing of Mus musculus
Unpublished (2001)
3 (bases 1 to 216959)
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 216959)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 216959)  
McPherson,J.D. and Waterston,R.H.  
Direct Submission  
Submitted (17-OCT-2002) Genome Sequencing Center, 4444 Forest Park  
Parkway, St. Louis, MO 63108, USA  
6 (bases 1 to 216959)  
Wilson,R.  
Direct Submission  
Submitted (08-NOV-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Oct 17, 2002 this sequence version replaced gi:21218578.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [submissions@watson.wustl.edu](mailto:submissions@watson.wustl.edu)  
----- Summary Statistics  
-----  
Center project name: M\_BA0016F06  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
Department of Genetics, Washington University, St. Louis MO. For  
additional information about the map position of this sequence, see  
<http://genome.wustl.edu>

#### SOURCE INFORMATION:

The RP23-23 BAC Library has been constructed by Kazutoyo Osegawa  
and Minako Tatenno in the laboratory of Pieter de Jong  
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or  
brain genomic DNA. The clone and detailed information can be  
obtained from Research Genetics, Inc. (<http://www.resgen.com>) or  
Pieter de Jong and coworkers at <http://www.chori.org>

#### NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES  
source

1. 216959  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="18"  
/map="18"  
/clone="RP23-16F6"  
/clone\_lib="RP23-23"  
1271..1406  
/rpt\_family="ERVK"  
1423..1535  
/rpt\_family="B4"  
1611..1806  
/rpt\_family="RMR1A"  
3133..3210  
/rpt\_family="ERV1"  
3256..3331  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region  
repeat\_region



QY 579 GTACTCGGTTGGTATACAGTCAMATATATTACTATTATTAAGATGCAATTAATCA 638  
Db 37508 TGTCTACTGTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 37567  
QY 639 TGGTACAGAAAAGTGTATCAATATGTTATAAGGATCTATGCCATCTGCTTCTGCTAGT 698  
Db 37568 TGTCTACTGTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 37627  
QY 699 TGATTGGAACGAGGCTCTTATGAAGTAATATTACTATGATGGAATCAGGAAATATTACAA 758  
Db 37628 TGTCTACTGTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 37687  
QY 759 TCTAATCTCAAGGTTCCGAAAAGCACTGGAAGTATTAACCTGTTAGAGGAAAATAATA 818  
Db 37688 TGTCTACTGTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 37747  
QY 819 TTTTCAGGATTACTATTCCTGGGCGAGCTACCAATATCTCCAAACCGGAAATCTCAAAATGG 878  
Db 37748 TGTCTACTGTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT 37807  
QY 879 AGCTAATGATGACTTTTTTTATTAAGGATAAATAACAATCACAGTCACATTTATACAG 934  
Db 37808 TGTCTACTGTCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTG 37863

RESULT 4

EX649545 171104 bp DNA linear HTG 16-OCT-2003  
LOCUS Danio rerio clone CH211-9N13, WORKING DRAFT SEQUENCE, 5 unordered  
DEFINITION pieces.  
ACCESSION EX649545  
VERSION EX649545.3 GI:37699896  
KEYWORDS HTG; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE Danio rerio (zebrafish)  
ORGANISM Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 171104)  
McLaren,S.  
Direct Submission  
Submitted (15-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Oct 16, 2003 this sequence version replaced gi:37051023.  
COMMENT  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zC9N13  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 170063 bases at least Q40  
Consensus quality: 170348 bases at least Q30  
Consensus quality: 170453 bases at least Q20  
Insert size: 170704; sum-of-contigs  
Insert size: 170285; 4.7% error; agarose-fp  
Quality coverage: 11.75x in Q20 bases; sum-of-contigs Quality  
coverage: 11.90x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 32930: contig of 32930 bp in length

FEATURES

source  
1..171104  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="CH211-9N13"  
/clone\_lib="CHORI-211"  
1..32930  
/note="assembly fragment:01076  
clone\_end:T7  
vector\_side:left"  
33031..63333  
/note="assembly fragment:00601  
fragment\_chain:1"  
63434..86315  
/note="assembly fragment:01493  
fragment\_chain:1"  
86416..117184  
/note="assembly fragment:00608  
fragment\_chain:2"  
117285..171104  
/note="assembly fragment:02797  
fragment\_chain:2  
clone\_end:SP6  
vector\_side:right"

ORIGIN

Query Match 5.9%; Score 89.6; DB 2; Length 171104;  
Best Local Similarity 44.2%; Pred. No. 7e-06;  
Matches 368; Conservative 0; Mismatches 464; Indels 0; Gaps 0;  
QY 189 TTTCGAATAAGATGAGCTTCTTATTAAATTCCTCAAGTAAGAAGCTGAGTATAAAGC 248  
Db 68152 TACAAATAAATACTATTAAAGTCTTTTGGCCTCTTTAGGCAGATTACTACTACTAC 68211  
QY 249 TTCAACTGATTTAAATTCCTTTTTCACCACTACTAATGAGGAGAACATATGTATAC 308  
Db 68212 TACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68271  
QY 309 TAAAAAGATGCTGCTGAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGTAA 368  
Db 68272 TACTATACTACTACAATCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68331  
QY 369 TACTACACAGTTTCCACTGTTTACTGAGTCAAAATATGATGCTGCTGAGGTTTAAATGT 428  
Db 68332 TACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68391  
QY 429 TTCCCAATATGATATATATTTGTTTCTAGACACTGTTTAAATATGAGGCTGTTAATGT 488  
Db 68392 TGCTACTACTACTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68451  
QY 489 TACATCTGTAACCTCCAAATGCTACTATTTCATGAAAGAAATATGATGCGACATGGGAGA 548  
Db 68452 TACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68511  
QY 549 TGGTGGTGGAAAACTGTAGATCAAAAAACGTAATCGGTTGGTGTATACAGTCAAAATATAC 608  
Db 68512 TACTGCTACTACTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 68571  
QY 609 TATTACTTATAGATGAGTCAATATCATGTCACAGAAAAGTGTATCAATATGTTAT 668  
Db 68572 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 68631  
QY 669 AAAGGATACTATGCCATCTGCTTCTGTTAGTTGATTTGAACGAAGGCTTTATGAGTAA 728







zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

## COMMENT

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

----- Project Information

Center project name: ZK46N5

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 111658 bases at least Q40

Consensus quality: 112600 bases at least Q30

Consensus quality: 113615 bases at least Q20

Insert size: 115507; sum-of-contigs

Insert size: 122756; 0.3% error; agarose-fp

Quality coverage: 7.13x in Q20 bases; sum-of-contigs Quality

coverage: 6.91x in Q20 bases; agarose-fp

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2116: contig of 2116 bp in length  
\* 2117 2216: gap of 100 bp  
\* 2217 8578: contig of 6362 bp in length  
\* 8579 8678: gap of 100 bp  
\* 8679 14990: contig of 6312 bp in length  
\* 14991 15090: gap of 100 bp  
\* 15091 26940: contig of 11850 bp in length  
\* 26941 27040: gap of 100 bp  
\* 27041 34475: contig of 7435 bp in length  
\* 34476 34575: gap of 100 bp  
\* 34576 41111: contig of 6536 bp in length  
\* 41112 41211: gap of 100 bp  
\* 41212 52851: contig of 11640 bp in length  
\* 52852 52951: gap of 100 bp  
\* 52952 58237: contig of 5286 bp in length  
\* 58238 58337: gap of 100 bp  
\* 58338 73639: contig of 15302 bp in length  
\* 73640 96363: contig of 22624 bp in length  
\* 96364 96463: gap of 100 bp  
\* 96464 99575: contig of 3112 bp in length  
\* 99576 104567: gap of 100 bp  
\* 104568 104667: contig of 4892 bp in length  
\* 104668 109346: contig of 100 bp  
\* 109347 109446: contig of 4679 bp in length  
\* 109447 116807: contig of 100 bp  
\* 116808 116807: contig of 7361 bp in length.

## FEATURES

## source

1. .116807  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-46N5"  
/clone\_lib="DanioKey"

## misc\_feature

1. .2116  
/note="assembly fragment:00043  
fragment chain:1"

## misc\_feature

2217. .8578  
/note="assembly fragment:00104  
fragment chain:1"

## misc\_feature

8679. .14990  
/note="assembly fragment:00322  
fragment chain:1"

## misc\_feature

15091. .26940  
/note="assembly fragment:00769  
fragment chain:1"

misc\_feature 27041. .34475  
/note="assembly fragment:00537  
fragment chain:1"  
misc\_feature 34576. .41111  
/note="assembly fragment:00392  
fragment chain:1"  
misc\_feature 41212. .52851  
/note="assembly fragment:00633  
fragment chain:2"  
misc\_feature 52952. .58237  
/note="assembly fragment:00201  
fragment chain:2"  
misc\_feature 58338. .73639  
/note="assembly fragment:00925  
fragment chain:2"  
misc\_feature 73740. .96363  
/note="assembly fragment:01125  
fragment chain:2"  
misc\_feature 96464. .99575  
/note="assembly fragment:00066.0"  
misc\_feature 99676. .104567  
/note="assembly fragment:00151"  
misc\_feature 104668. .109346  
/note="assembly fragment:00258"  
misc\_feature 109447. .116807  
/note="assembly fragment:00464"

## ORIGIN

Query Match 5.7%; Score 86; DB 2; Length 116807;  
Best Local Similarity 44.7%; Pred.No.3.1e-05;  
Matches 335; Conservative 0; Mismatches 415; Indels 0; Gaps 0;  
QY 280 ACTACTTAATGGAGGAGCAATATGTAATACTAAAGAAAGTACGCGTCAGCAAAATGAGATT 339  
Db 26253 AATA 26194  
QY 340 GGCACATGGCTAAATCTATATCAGCTAATACTACACCACTGTTCCAGTGTCTACTGAGTCA 399  
Db 26193 GATA 26134  
QY 400 AATAATGATGGTACTGAGGTATTATGTTCCCAATATGGATATATATTTCTTAGC 459  
Db 26133 AATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 26074  
QY 460 ACTGTTAATAATGGAGCTGTATATGTTTAACTCTGTAACCTCAAATGCTACTATTTCAT 519  
Db 26073 AATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 26014  
QY 520 GAAAAGATCTGATCGGACATGGGAGATGGTGGGAAAACCTGTAGATCAAAAACG 579  
Db 26013 AATA 25954  
QY 580 TACTCGTTGGTGATACAGTCAAAATATATCTATTCTATAAGAAATGCAGTCAATATCAT 639  
Db 25953 AATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25894  
QY 640 GGTACAGAAAAAGTGTATCAATATGTTTAAAGGATACCTATGCCATCTCTCTCTGTAGTT 699  
Db 25893 AATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 25834  
QY 700 GATTTGAACGAGGGCTTTATGAAGTACTATTCTAGTGGATCAGGGGAATATTACAAT 759  
Db 25833 AATGATA 25774  
QY 760 CTAACCTCAAGGTTCCGAAAAGCAACTGGGAAGTATAACCTGTAGAGAAAATAATAATAATA 819  
Db 25773 AATA 25714  
QY 820 TTCACGATTACTTTCGTTGGGAGCTACCAATCTCTCAACCGGAAAATCTCAAAATGGA 879  
Db 25713 GATA 25654  
QY 880 GCTAATGATGACTTTTTTTTATAAGGAATAATAACAATCACAGTCACTTATACAGGAGTA 939







SOURCE ORGANISM	REFERENCE AUTHORS TITLE JOURNAL	COMMENT
Danio rerio (zebrafish)		
Danio rerio		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
1 (bases 1 to 94534)		
Almeida, J.		
Direct Submission		
Submitted (16-SEP-2003)		
Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk		
Clone requests: clonerequest@sanger.ac.uk		
On Sep 16, 2003 this sequence version replaced gi:27848021.		
----- Genome Center		
Center: Wellcome Trust Sanger Institute		
Center code: SC		
Web site: http://www.sanger.ac.uk		
Contact: zfish-help@sanger.ac.uk		
-----		
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a VAC.		
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep		
CH211-137A2 is from a CHORI-211 BAC library VECTOR: pTARBAC2.1		
Clone-derived zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.		
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhiyong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml.		
Location/Qualifiers		
1. 94534		
/organism="Danio rerio"		
/mol_type="genomic DNA"		
/db_xref="taxon:7955"		
/clone="CH211-137A2"		
/clone_lib="CHORI-211"		
ORIGIN		
Query Match	5.5%;	Score 82.4; DB 5; Length 94534;
Best Local Similarity	44.5%;	Pred. No. 0.00013;
Matches 326; Conservative	0;	Mismatches 406; Indels 0; Gaps 0;
QY	103	ATTACAGTTCGAAGTACTCAAAAAGGCGCAACCTATAAAGCATATAAGCTTTTGTGATCA 162
Db	30450	AATATAGTTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 30391
QY	163	GAATAGATAATGCAATGTATCTGATTCGAATAAAGATGGAGCTTCTTTATTTATTCCT 222
Db	30390	AATAATATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT 30331
QY	223	CAAGGTAAGAGCTGAGTATAAAGCTTCAACTGATTTTAATTTCTTTTTCAGCAACT 282
Db	30330	AATAATAATAATGATGATAATAAAGATAATAATAATAATAATAATAATAATAATAATA 30271

283 ACTAATGGAGGAGAACATATGTAACTAAAAAGATACCTGCGTCAGCAAAATGAGATTCCG 342  
30270 AATAATGATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 30211  
343 ACATGGGCTAATCTATACAGCTAATCTACACCAAGTTTCCACCTGTTACTGAGTCAAAAT 402  
30210 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 30151  
403 AATGATGGTACTGAGGCTTAATTAATGTTTCCCAATATGGAATATTAATGTTTCTAGCACT 462  
30150 AATAATGGTGATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAAC 30091  
463 GTTAATAATGGAGCTGTAATATATGTTTACATCTGTAATCTCCAAATGCTACTACTTATCATGAA 522  
30090 AGTAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAGT 30031  
523 AAGAATACTGATCGCATGGGAGATGGTGGGAAAACTGTAGATCAAAAAACGTAC 582  
30030 AATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 29971  
583 TCGGTTGGTGATACAGTCAAAATATATCTATTAAGAAATGCGAGTCAATATCATGCT 642  
29970 AATAATAATGGTGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATGATGATAAT 29911  
643 ACAGAAAAAGTGTATCAATATGTTATAAAGGATACCTATGCACTCTGTTCTGTAGTTGAT 702  
29910 AAT 29851  
703 TTGAACGAAGGCTTATGAACTAATCTATTAATGATGATGATGATGATGATGATGATGAT 762  
29850 AGTAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAGCGAT 29791  
763 ACTCAAGTTCCGAAAAAGCACTGGGAAGTATAACCTGTTAGAGGAATAATAATAATTTTC 822  
29790 AATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATGATAAT 29731  
823 ACGATTACTATT 834  
29730 AATAATTAATAAT 29719

RESULT 12

PFMAL13\_07

WPCOMMENT

Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509

Fragment Name	Begin	End
PFMAL13_00	1	110000
PFMAL13_01	100001	210000
PFMAL13_02	200001	310000
PFMAL13_03	300001	410000
PFMAL13_04	400001	510000
PFMAL13_05	500001	610000
PFMAL13_06	600001	710000
PFMAL13_07	700001	810000
PFMAL13_08	800001	910000
PFMAL13_09	900001	1010000
PFMAL13_10	1000001	1110000
PFMAL13_11	1100001	1210000
PFMAL13_12	1200001	1310000
PFMAL13_13	1300001	1410000
PFMAL13_14	1400001	1510000
PFMAL13_15	1500001	1610000
PFMAL13_16	1600001	1710000
PFMAL13_17	1700001	1810000
PFMAL13_18	1800001	1910000
PFMAL13_19	1900001	2010000
PFMAL13_20	2000001	2110000
PFMAL13_21	2100001	2210000
PFMAL13_22	2200001	2310000
PFMAL13_23	2300001	2410000
PFMAL13_24	2400001	2510000
PFMAL13_25	2500001	2610000
PFMAL13_26	2600001	2710000

PFMAL13\_27 2732359  
Continuation (8 of 28) of PFMAL13 from base 700001 (AL844509 Plasmodium falciparum 3D7 cl

Query Match 5.3%; Score 80.6; DB 2; Length 110000;  
Best Local Similarity 44.8%; Pred. No. 0.00024;  
Matches 308; Conservative 0; Mismatches 379; Indels 0; Gaps 0;

137 ATAAGCATATAAAGTTTGTGATGCAGAAATAGATAATGCAATGTATCTGATTCGAATA 196  
43055 ATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATGATAATAATA 43114  
197 AAGATGAGGCTTCTTAATTAATTCCTCAGGTAAGAGCTGAGTATAAGCTTCAACTG 256  
43115 ATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATGATAATAATA 43174  
257 ATTTTAATTCCTTTTACGCAAACTACTAATGGAGGAGAACATATGTAACATAAAAAAG 316  
43175 ATAATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATGATAATA 43234  
317 ATACTGGCTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAACTACAC 376  
43235 ATAATAATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATG 43294  
377 CAGTTTCCACTGTTACTGAGTCAAAATATGATGATGATGAGGTTTATTAATGTTTCCCAAT 436  
43295 ATAATAATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATG 43354  
437 ATGGATATTAATTAATGTTTCTAGCACTGTTAATTAATGAGCTGTAATTAATGGTTACATCTG 496  
43355 ATGATAATAATAATGATGATGATAATAATAATGATGATGATAATAATAATAATGATG 43414  
497 TAACTCCAAATGCTTACTTATTCATGAAAGATATCTGATGCGACATGGGAGATGGTGGT 556  
43415 ATGATGATAATAATAATGATGATGATAATAATAATGATGATGATAATAATAATGATG 43474  
557 GAAAACTGTAGTCAAAAAACGTAATCTGGTGGTGATACAGTCAAAATATATCTTACTTACT 616  
43475 ATGATGATGATAATAATAATGATGATGATAATAATAATGATGATGATAATAATAATA 43534  
617 ATAAGAATGCACTCAATTAATCATGGTACAGAAAAGTGATCAATATGTTATAAGGATA 676  
43535 ATGATGATGATGATAATAATAATGATGATGATAATAATAATGATGATAATAATAATCAT 43594  
677 CTATGCCATCTCTCTCTAGTTGATTTGAACGAGGCTCTTATGAACTAACTATTACTG 736  
43595 ATAATAATCATTAATAATAATTAATTAATAATAATAATAATAATAATAATAATAATAAG 43654  
737 ATGGATCAGGGAATATTACAACTCTTAACAGGTTCCGAAAAGCACTGGGAAGTATA 796  
43655 AAGAAGTATCAATTAATAATATCTGAACCCCCCAATGCTATAGGAACATAATATGGTAAT 43714  
797 ACTGTTAGAGGAATAATAATAATTAATTTCA 823  
43715 ATATTTCCAGGATGAAGATAATAATTAACA 43741

RESULT 13

BX537105

LOCUS

DEFINITION

Accession

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission

Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

BX537105 225581 bp DNA linear HTG 06-AUG-2003  
Danio rerio clone DKEY-3L24, 18 unordered pieces.  
BX537105  
BX537105.2 GI:31559379  
HTG; HTGS PHASE1; HTGS CANCELLED.  
Danio rerio (zebrafish)

Danio rerio  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 225581)

Burton, J.  
Direct Submission  
Submitted (05-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Jun 9, 2003 this sequence version replaced gi:31441851.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zK3L24  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 217427 bases at least Q40  
Consensus quality: 219130 bases at least Q30  
Consensus quality: 220818 bases at least Q20  
Insert size: 223881; sum-of-contigs  
Insert size: 219000; 7.3% error; agarose-fp  
Quality coverage: 6.82x in Q20 bases; sum-of-contigs Quality  
coverage: 7.43x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 20602: contig of 20602 bp in length  
\* 20603 20702: gap of 100 bp  
\* 20703 57831: contig of 37129 bp in length  
\* 57832 57931: gap of 100 bp  
\* 57932 65164: contig of 7233 bp in length  
\* 65165 65264: gap of 100 bp  
\* 65265 71392: contig of 6128 bp in length  
\* 71393 71492: gap of 100 bp  
\* 71493 93677: contig of 22185 bp in length  
\* 93678 93777: gap of 100 bp  
\* 93778 101598: contig of 7821 bp in length  
\* 101599 128689: gap of 100 bp  
\* 128690 128789: contig of 26991 bp in length  
\* 128790 138632: contig of 9843 bp in length  
\* 138633 138732: gap of 100 bp  
\* 138733 144193: contig of 5461 bp in length  
\* 144194 144293: gap of 100 bp  
\* 144294 155818: contig of 11525 bp in length  
\* 155819 168354: gap of 100 bp  
\* 168355 168454: gap of 100 bp  
\* 168455 171301: contig of 2847 bp in length  
\* 171302 171401: gap of 100 bp  
\* 171402 180364: contig of 8963 bp in length  
\* 180365 180464: gap of 100 bp  
\* 180465 188604: contig of 8140 bp in length  
\* 188605 188705: gap of 100 bp  
\* 188705 192855: contig of 4151 bp in length  
\* 192856 192955: gap of 100 bp  
\* 192956 201754: contig of 8799 bp in length  
\* 201755 201855: gap of 100 bp  
\* 201855 211372: contig of 9518 bp in length  
\* 211373 211472: gap of 100 bp  
\* 211473 225581: contig of 14109 bp in length.

FEATURES

Source

1..225581  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-3L24"  
/clone\_lib="DanioKey"  
1..20502  
/note="assembly\_fragment:01463  
fragment\_chain:1"

misc\_feature

misc\_feature 20703..57831  
/note="assembly\_fragment:01892  
fragment\_chain:1"  
misc\_feature 57932..65164  
/note="assembly\_fragment:02319  
fragment\_chain:1"  
misc\_feature 65265..71392  
/note="assembly\_fragment:02166  
fragment\_chain:1"  
misc\_feature 71493..93677  
/note="assembly\_fragment:01041  
fragment\_chain:1"  
misc\_feature 93778..101598  
/note="assembly\_fragment:02235  
fragment\_chain:1"  
misc\_feature 101699..128689  
/note="assembly\_fragment:01236  
fragment\_chain:1"  
misc\_feature 128790..138632  
/note="assembly\_fragment:02758  
fragment\_chain:1"  
misc\_feature 138733..144193  
/note="assembly\_fragment:02245  
fragment\_chain:1"  
misc\_feature 144294..155818  
/note="assembly\_fragment:01852  
fragment\_chain:1"  
misc\_feature 155919..168354  
/note="assembly\_fragment:01722  
fragment\_chain:1"  
misc\_feature 168455..171301  
/note="assembly\_fragment:00772  
fragment\_chain:2"  
misc\_feature 171402..180364  
/note="assembly\_fragment:02453  
fragment\_chain:2"  
misc\_feature 180465..188604  
/note="assembly\_fragment:00524  
fragment\_chain:2"  
misc\_feature 188705..192855  
/note="assembly\_fragment:02841  
fragment\_chain:2"  
misc\_feature 192956..201754  
/note="assembly\_fragment:00644  
fragment\_chain:2"  
misc\_feature 201855..211372  
/note="assembly\_fragment:02536  
fragment\_chain:2"  
misc\_feature 211473..225581  
/note="assembly\_fragment:01373  
fragment\_chain:2"

ORIGIN

Query Match 5.3%; Score 80.4; DB 2; Length 225581;  
Best Local Similarity 45.3%; Pred. No. 0.00021;  
Matches 291; Conservative 0; Mismatches 351; Indels 0; Gaps 0;  
QY 261 TAAATCTCTTTTACGCAACTACTAATGAGGGGAGACATATGTAACATAAAGATAC 320  
DB 151722 TACTTCTACTACTACTACTACTTATAATAATAATAATAATAATAATAATA 151781  
QY 321 TGGGTACGCAATGAGATTGGACATGGGCTAAATCTATATCAGTAACTACTACACAGT 380  
DB 151782 TAATAATAATAATGATATAATAATAATAATAATAATAATAATAATAATAATA 151841  
QY 381 TTCACCTGTTACTAGTCAATAATAATGATGGTACTGAGGTATTATGTTCCCATATGG 440  
DB 151842 TAATAATAATAATGATATAATAATAATAATAATAATAATAATAATAATAATA 151901  
QY 441 ATATTATTATGTTCTAGCACTGTTAATAATAATGAGGAGCTGTAATTATGTTTACATCTGTAAC 500  
DB 151902 TAATGATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 151961

501	TCCAAATGCTACTATTTCATGAAAGAAATACTGATCGGACATGGGAGATGGTGGTGA	560
151962	TATTGATAATAATGATAATAATAATAATGATAATAATAATAATAATAATAATAATA	152021
561	AACTGTAGATCAAAAACGCTACTCGGTTGGTGATACAGTCAAAATATACTATTACTTATA	620
152022	TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	152081
621	GAATGCAGTCAATTTATCATGGTACAGAAAAAGTGTATCAATATGTTATAAAGGATACTAT	680
152082	TAATGATAATAATAATGATAATAATAATAATAATAATGATAATAATAATAATAATA	152141
681	GCATCTGCTTCTGTAGTTGATTTCGAACGAGGGCTTATGAAGTAACTATTACTGATGG	740
152142	TAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	152201
741	ATCAGGGAATATTACAACTCTAACTCAAGGTTTCGGAAAAAGCACTGGGAAGTATAACCT	800
152202	TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	152261
801	GTTAGAGGAAATAATAATTTCCAGATTACTATTCGGTGGGAGCTACCAATATCTCCAC	860
152262	TAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	152321
861	CGGAATATCTCAAAATGGAGCTTAATGATGACGACCTTTTTTTTATAA	902
152322	TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	152363

RESULT 14

AF397467/c

LOCUS

AF397467 10115 bp DNA linear VRT 27-NOV-2001

DEFINITION

Ictalurus punctatus novel immune-type receptor 2 (NITR2) and novel immune-type receptor 4 (NITR4) genes, complete cds.

ACCESSION

AF397467

VERSION

AF397467.1 GI:17105096

KEYWORDS

Ictalurus punctatus (channel catfish)

SOURCE

Ictalurus punctatus

ORGANISM

Ictalurus punctatus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes; Ictaluridae; Ictalurus.

AUTHORS

1 (bases 1 to 10115)  
Hawke, N.A., Yoder, J.A., Haire, R.N., Mueller, M.G., Litman, R.T., Miraciac, A.L., Stuge, T., Shen, L., Miller, N. and Litman, G.W.

TITLE

Extraordinary variation in a diversified family of immune-type receptor genes

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 98 (24), 13832-13837 (2001)

MEDLINE

21574523

PUBMED

11698645

REFERENCE

2 (bases 1 to 10115)  
Hawke, N.A., Yoder, J.A., Haire, R.N., Mueller, M.G., Litman, R.T., Miraciac, A.L., Stuge, T., Miller, N. and Litman, G.W.

AUTHORS

Direct Submission

TITLE

Submitted (09-JUL-2001) Department of Pediatrics, University of South Florida Children's Research Institute, 140 Seventh Avenue South, St. Petersburg, FL 33701, USA

JOURNAL

Location/Qualifiers

FEATURES

1..10115

source

/organism="Ictalurus punctatus"

/mol\_type="genomic DNA"

/db\_xref="taxon:7998"

<2836..>4401

/gene="NITR2"

join(<2836..2875,3025..3381,3759..4112,4226..>4401)

/gene="NITR2"

/product="novel immune-type receptor 2"

join(2836..2875,3025..3381,3759..4112,4226..4401)

/gene="NITR2"

/note="IpNITR2"

/codon\_start=1

/product="novel immune-type receptor 2"

/protein\_id="AAL35555.1"

RESULT 15

/db\_xref="GI:17105097"

/translation="MLTACFIILLCPHLVKQTDFRNSDSLVAKEGDSVNISCIVE

SDMAFSPYKNLQCKPLISNFYKDKKATFHEFKNARFSPWNEKSKNLEIKG

LQSDSATYFCGSAHSNIIVERGEGTELVAQSHLSVLQOPVHELAPHSTLHCT

VITDRCAGHSVYVFRHNSGRCKSPGVIIYTHGDSNGRCKKPKRAGSLTRMCVSLPKTN

LSTSDYGTGTHCAVACGQLFGNGTKDKMITGLGVFENILQTLVLFSIVRSVVLCCV

ILIM1YPCSKKAKCSVRRTTKGKNLAQLEK"

<7826..>9190

/gene="NITR4"

join(<7826..7865,8018..8368,8631..8975,9084..>9190)

/gene="NITR4"

/product="novel immune-type receptor 4"

join(7826..7865,8018..8368,8631..8975,9084..9190)

/gene="NITR4"

/note="IpNITR4"

/codon\_start=1

/product="novel immune-type receptor 4"

/protein\_id="AAL35556.1"

/db\_xref="GI:17105098"

/translation="MILYFFLLCTSDCLYIEEVMQEPLETTAEDGGSVKLTCLYTK

QRTTIIWLROKLGEPYVAVSYQLQPAQYVNI FEKNVFNARI GPSSFNLSINIK

LSDSATYCAVATFLYEITFGQGTVLIVKDSLKRTVQOEKVMPEVPGDSVQCITIV

TDRCAGHNHWFHSGSDSGIIVTQSGSGCKESFNASSSOTCVYSLPKNLRV

SDAGTYICVAVVACGEILCGSEAKQNIKDVHFQIEVLVLLSIIRSAVLLFIICLLCT

YIGH"

ORIGIN

Query Match 5.3%; Score 80; DB 5; Length 10115;

Best Local Similarity 47.5%; Pred. No. 0.00064;

Matches 270; Conservative 0; Mismatches 295; Indels 3; Gaps 1;

QY 171 TAATGCAAAATGTAATCTGATTCGAATAAAGATGGAGCTTCTTATTATTAATCTCAAGTAA 230

Db 6797 TACTACTACTATTAATCTGCTACTGCTGTTACTACTGCTGCTACTACTACTACTACTAC 6738

QY 231 AGAAGCTGAGTATAAGCTTCAACTGATTTAACTTCTCTTTTACGACCAACTACTAATGG 290

Db 6737 TGCTACTATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6678

QY 291 AGGGAGAACATATGTAACATAAAGATACTGCTGCTGAGCAATGAGATGGACATGGGC 350

Db 6677 TACTAGTACTACTGCTACTACTATTAATCTGCTGCTACTACTACTACTACTACTACTAC 6618

QY 351 TAAATCTATATCAGTAATATCTACACAGTTCCTCCACTGTTTACGAGTCAAAATATGATG- 409

Db 6617 TATTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 6558

QY 410 --GTACTGAGGTTTAAATATGTTTCCCAATATGGATATTATATTATGTTTCTAGACATGTTAA 467

Db 6557 TACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 6498

QY 468 TAATGGAGCTGTAATATGTTGTTACATCTGTAACCTCAATCTACTACTACTACTACTAAAGAA 527

Db 6497 TACTGCTGCTACTACTAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 6438

QY 528 TACTGATCGACATGGGAGATGGTGGGAAAACTGTAGATCAAAAAACGTACTCGGT 587

Db 6437 TGCTGCTACTACTAGTACTGCTGCTATTACTACTGCTGCTACTACTACTACTACTACTACTAT 6378

QY 588 TGGTGATACAGTCAAAATATATCTATTATTAAGATGCAAGTCAATATATCTGCTACAGA 647

Db 6377 GATTACTACTACTACTCTCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 6318

QY 648 AAAAGTGTATCAATATGTTTATAAGGATACTATGCAATCTGCTCTGTTAGTTGAA 707

Db 6317 TACTACTAATACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTCTAC 6258

QY 708 CGAAGGCTCTTATGAAGTAACTATTACT 735

Db 6257 TACTACTAATACTAGCATTTACTACTACT 6230

```

AY333070/c
LOCUS       AY333070             308092 bp    DNA    linear    INV 03-SEP-2003
DEFINITION  Drosophila virilis antennapedia (Antp), CG10013 (CG10013), CG31217
            (CG31217), and ultrabithorax (Ubx) genes, complete cds.
ACCESSION   AY333070
VERSION     AY333070.1  GI:34398397
KEYWORDS    Drosophila virilis
SOURCE      Drosophila virilis
ORGANISM   Drosophila virilis
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 308092)
AUTHORS     Lewis,E.B., Pfeiffer,B.D., Mathog,D.R. and Celniker,S.E.
TITLE       Evolution of the homeobox complex in the Diptera
JOURNAL     Curr. Biol. 13 (15), R587-R588 (2003)
MEDLINE     22788672
PUBMED      12906807
REFERENCE   2 (bases 1 to 308092)
AUTHORS     Pfeiffer,B.D., Lewis,E.B., Park,S., Wan,K. and Celniker,S.E.
TITLE       Direct Submission
JOURNAL     Submitted (30-JUN-2003) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley,
            CA 94720, USA
FEATURES    Location/Qualifiers
             1..308092
             /organism="Drosophila virilis"
             /mol_type="genomic DNA"
             /db_xref="taxon:7244"
             /chromosome="2"
             /map="24E"
             complement (<2176..>14870)
             /gene="Antp"
             complement (join(<2176..2426,13726..13960,14082..14120,
             14253..>14870))
             /gene="Antp"
             /product="antennapedia"
             /note="maximized conservation of amino acids was used
             rather than adherence to consensus splice sites"
             complement (join(2176..2426,13726..13960,14082..14120,
             14253..14870))
             /gene="Antp"
             /note="transcription factor; similar to the Drosophila
             virilis antennapedia protein encoded by GenBank Accession
             Number W95825"
             /codon_start=1
             /product="antennapedia"
             /protein_id="AA067266.1"
             /db_xref="GI:34398398"
             /translation="WTMTSTNCESMTSYFTNSYMGADMHGHYPGNGVTDLDAAQMH
             YSQNPQGNMPPRPYPDRMPYNGQMDQQOQHQGRPSRPSQSGVGMPPAQ
             TNGQLVSAQQOQTQQOQAOQQOQAPLQOQHPVQTQVTHPQQOQPVVYAS
             CKLQAVGLGVQSGSPPLVDQMGHMHNAQMTLPHHMGHPQALQGVTDVGDVT
             EVQHNHNGMYQQOQTGVPPVAPQAMHFGAGQPPQHQHGPQHTPPSQNPSS
             QSGMPSPLPMPWQFGKQGRQRTYTRYQLEKEKFEHFNRYLTRRRRIEIAH
             ALCLTERQIKIWFQNRMKWKKENTKGEPGSGGDEITPNSPQ"
             148735..149734
             <169031..>170377
             /gene="CG10013"
             <169031..>170377
             /db_xref="GI:34398399"
             /translation="WPLELIVISLMRTSKWINVLMNSPDEKARQOTRRIDVSANGI
             RIERDFEOTPLGKFEINTPEAGITVDKETNNATALKLSLVHINDAALDQOQVDV
             SIEDLHLLNESMACGLHSCNCLVPRYDCLIRMPVYTMEDIKYFCPSDKKPTS
             LAEGELLYALNTIVISPKILGKRVTCQGHITCGRCLQVGSGLGKVKATQILYVDTLH
             VAPRQGSQVLPGEGLQFGLQATASQLMLHLQDAVPISEKTRFLKTVRPDGLH
             YMLLVDTSSQHLHRSKLSLIEDLVSSSLKYTDDDTNEDSSSESSESDCTSDSLD
             WREGTDBDEGPKAKRRRAETKPVHVGKRGYGRGCRINYHIFSDSELAANHEMK
             ARRTGPMLRISYMMMDLRLNELLVATLERIAPDEKEGHIYIYISPVEDM"
             complement (<196716..>201570)
             /gene="CG31217"
             complement (join(<196716..197220,197761..198290,
             198388..198868,198935..199244,201396..>201570))
             /gene="CG31217"
             /product="CG31217"
             /note="maximized conservation of amino acids was used
             rather than adherence to consensus splice sites"
             complement (join(196716..197220,197761..198290,
             198388..198868,198935..199244,201396..>201570))
             /gene="CG31217"
             /note="serine protease; trypsin family"
             /codon_start=1
             /product="CG31217"
             /protein_id="AAQ67268.1"
             /db_xref="GI:34398400"
             /translation="MEGACWQHIGLNMSTGQADRLPATNVPHQCFGLTEGARL
             CALITLYPIASIHRAVYGACPTKFCNDGSCISKFDACDGVKRNCPDGSDEGTQCV
             SHRKCKPYQCSYGACVIGTAPCNKGCADSDETVLRCGTDGDDILEPDKLQOQN
             CQSNFEKCPGICIDKSKYLCDDGDDCGDGSYDEAVKFCGHIKICGVPARFKASGGCI
             SNQITCKGKDCPDGSEAPLLCNTTCKPSQPLVKPTNQMGCLPVGDEQPLKDKHR
             GYLTPTPIRATVYFSCNPGVLEGAESSHCANGKWSLPIVPRCVSSSGHKNLNGY
             STRATCYNGQQNQGRYHPPEVTEVFCVATGFTLTTLTPDKLPSGHWSRER
             CDQRCGEIATPIKQFSAFYSVNNTVPMWGLVYLNERNDFVCGGTLLTLPDIVIT
             AAHCVFQSVSBOTSLATFRVAAKLYGVNDVTSDEKRDVKSIKVAAPRYKGRDNF
             LNDLALVLEPYOLSNVIRPVCPDVALKESVNNVRGQFAGWFKDKRELQFVP
             AESKINAVCGTSLRDIGAKFCMTYQKSLACHGDSGGGPTAQRETQFSRVNTHYL
             YGVISSAPNAGCAQSLLTNTIQTFFEDMINDAIRESEFSGS"
             complement (<203957..>295030)
             /gene="Ubx"
             complement (join(<203957..204261,265785..265835,
             284419..284469,294274..>295030))
             /gene="Ubx"
             /product="ultrabithorax"
             complement (join(<203957..204261,265785..265835,
             284419..284469,294274..>295030))
             /gene="Ubx"
             /note="transcription factor"
             /codon_start=1
             /product="ultrabithorax"
             /protein_id="AAQ67269.1"
             /db_xref="GI:34398401"
             /translation="MNSYFEQASGFYGHPHQATGWTGTHDQTTASAAAAAYRGF
             PLSIGMSPYANHILQRTQDSPYDASITAACNKYGDAGSAVKODCLNIKADAVNGYK
             DIWNTGANGSGGGGGGAGTAGTNSGNAPNAANGONNAGGMPVPSPACTPDSRV
             GGYLDTSQSGSPVSHRGGSGAGNGTAGVFPQNTASGVGGVGTAGTANACTISGAA
             AAQTAASSLSLQASNNHTFYPMMAJAGSTADPSKISDITQYGGISTDMGKRYSES
             LAGSLLPDLWLTNGLRGRQYTRYQYQLEKEKFEHFNRYLTRRRRIEMAHLCVTER
             QIKIWFQNRMRMLKKEIQAIKELNEQBKQAQAKAAAAAQAQVGGHLDQN"
ORIGIN
Query Match      5.3%; Score 79.8; DB 3; Length 308092;
Best Local Similarity 43.1%; Pred. No. 0.00024;
Matches 439; Conservative 0; Mismatches 577; Indels 3; Gaps 1;
QY      65  CAGTTACGCCGATAGCTTTTCCGCTGAGACAGGACACATTACAGTTCAGATCTCAAA 124
DB      167307  CAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATA 167248
QY      125  AAGCGCAACCTATAAAGCATATAAAGTTTTTGTATGACAGAAATAGATAATGCAATGTAT 184
DB      167247  CAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATA 167188
QY      185  CTGATTCGAATAAGATGGAGCTTCTTTATTTTAAATTCCTCAAGGTTAAAGAGCTGATATA 244
DB      167187  CAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 167131
QY      245  AAGCTTCAACTGATTTTAAATCTCTTTTACGACAACTACTAATGAGGGGGAACATATG 304
DB      167130  ATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAG 167071

```













SOURCE	Danio rerio (zebrafish)	QY	395	AGTCAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGGATATTATTATGTTT	454
ORGANISM	Danio rerio	Db	79126	CTATTACTAATCAAAATACTACTACTATTATTACTACTGCTAAAAAATACTACTCTG	79185
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.	QY	455	CTAGCAGCTGTTTAATAATGAGAGCTGTAAATGTTTACATCTGTAATCCAAATGCTACTTA	514
AUTHORS	1 (bases 1 to 200634)	Db	79186	CTATTACTACTACTAATCAAACTACTACTACTAATCAAAATACTGCTACTACTTA	79245
TITLE	Mashreghi-Mohammadi, M.	QY	515	TTTCATGAAAGAATACTGATCGGAGATGGGGAGATGGTGGGAAATACTGTAGATCAAA	574
JOURNAL	Submitted (03-JUL-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk	Db	79246	CTACTACTAATCAAACTACTACTACTAATCAAACTACTACTACTACTACTAATA	79305
COMMENT	On Jul 1, 2003 this sequence version replaced gi:3216885. ----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: <a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a> Contact: zfish-help@sanger.ac.uk -----	QY	575	AAACGTACTCGTGGTGATACAGTCAATATCTATTATCTATTATTAAGAAATGAGTCAATT	634
		Db	79306	CTAATCAAAATACTGCTACTACTACTAATCAAACTACTACTACTACTACTACTAAT	79365
		QY	635	ATCATGGTACAGAAAGAGTGTATCAATATGTTTATAAGGATCTATGTCATCTGCTCTG	694
		Db	79366	AACTACTACTACTAATCAAACTACTACTACTAATCAAACTACTACTACTACTACTA	79425
		QY	695	TAGTTGATTGAACGAGAGGGTCTTTATGAAGTAATCTATTACTGTGATGAGTACAGGAATATTA	754
		Db	79426	CTATTAGTACTACTACTAGCACTATTA---TTACAACCTACTAATCAAACTACTACTACTA	79482
		QY	755	CAACTCTAATCAAGTTCGGGAAAGCACTGGGAAGTATAACCTGTTAGAGGAAATA	814
		Db	79483	CTACTAATATTAATAATAATCAAACTACTACTACCCTATTACTACTAAACCAACTACTA	79542
		QY	815	ATAATTTCCAGCATCTACTATTCGGTGGGAGCTACCAATCTCCCAACCGGAATACTCAAA	874
		Db	79543	ACATTACTACTACTACTACTCAAACTACTACTACTAATCAAACTACTACTACTACTA	79602
		QY	875	ATGAGCTTAATGATGACATTTTTTTTATGAAGGAATAATACTACACAGTCACTTATACAG	934
		Db	79603	CTATTACTACTAATCAAACTACTACTACTACTAATCAAACTACTACTACTACTACTAC	79662
		QY	935	GAGTATTAAGAGTGGAGCTAAACAGGTTCCAGCTGATTTACCAGAAATCAAAACATTTG	994
		Db	79663	TACCCTATTACTACTACTACCAAACTA-----CTAACACTACTACTACTACTACTACTA	79715
		QY	995	CGACCATCAACCCCAATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG	1054
		Db	79716	CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTC	79775
		QY	1055	GTCAAATTTACTATATAAAAAATTTGATGTTTCCACAAAAGCTTC	1097
		Db	79776	CTACTATTACTACCACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	79818
RESULT 21	AL935062	176010 bp	DNA	linear	VRT 30-AUG-2003
LOCUS	Zebrafish DNA sequence from clone CH211-231P23 in linkage group 9,				
DEFINITION	complete sequence.				
ACCESSION	AL935062				
VERSION	AL935062.13	GI:34368574			
KEYWORDS	HTG.				
SOURCE	Danio rerio (zebrafish)				
ORGANISM	Danio rerio				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.				
REFERENCE	1 (bases 1 to 176010)				
AUTHORS	Tromans, A.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-AUG-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk				
COMMENT	On Aug 30, 2003 this sequence version replaced gi:32127703. ----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: <a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a>				

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep)

Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see [http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml)

is from a Zebrafish BAC library

VECTOR: pIndicoBAC-5

Location/Qualifiers

1..200634

/organism="Danio rerio"

/mol\_type="genomic DNA"

/db\_xref="taxon:7955"

/clone="DKEY-12H12"

/clone\_lib="DanioKey"

Query Match 5.1%; Score 77.4; DB 5; Length 200634;  
Best Local Similarity 45.4%; Pred. No. 0.00067;  
Matches 401; Conservative 0; Mismatches 471; Indels 11; Gaps 3;

QY 215 TAATTCCTCAAGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTCTTTTCTTTT 274

Db 78947 TAAAGATCACAATAAATGCTATTACTACTATCAATGAATCACTACTACTGATA 79006

QY 275 CGACAACTACTAATGGAGGAGACATATGTAACATAAAAGATAGTGGTCAGCAATG 334

Db 79007 CTACTACTACTAATCAAACTACTACTACTAATTAATAAACTACTACTACTACTA 79066

QY 335 AGATTCCGATGGGGTAAATCTATATCACTAGTAAATCACTACACAGTTTCCACTGTTACTG 394

Db 79067 CTACTGCTAAAAA-AATATTACTATTACTCTCTAATCAAACTACTACTACTACTA 79125





```
/note="MLL-like"
/pt_type=OTHER
4455..4733
/gene="Furu-I103"
/number=3
4734..4902
/gene="Furu-I103"
/number=3
4903..5229
/gene="Furu-I103"
/number=4
5230..5464
/gene="Furu-I103"
/number=4
5465..5569
/gene="Furu-I103"
/number=5
5570..5714
/gene="Furu-I103"
/number=5
5715..5729
/gene="Furu-I103"
/number=6
5730..5843
/gene="Furu-I103"
/number=6
5844..5893
/gene="Furu-I103"
/number=7
complement(7521..9972)
/gene="psmb8"
complement(join(7521..7609,7917..8112,8861..8990,
9167..9278,9460..9601,9823..9972))
/gene="psmb8"
complement(join(7521..7609,7917..8112,8861..8990,
9167..9278,9460..9601,9823..9972))
/gene="psmb8"
/codon_start=1
/product="low molecular mass polypeptide subunit PSMB8"
/protein_id="CAC13117.1"
/db_xref="GI:10803370"
/db_xref="GOA:Q9DES2"
/db_xref="SPTREMBL:Q9DES2"
/translation="MALFDVSDTRYSLRRLQPSGQRLKERTTHYSFGLLKQFFA
VPLGVDPSMEIKSCNNEAGVDIELDGTTLAPFKKHGIVAVDSASAGKYISSDA
KXVEINSYLLGTWSSAADCNTWERRLLAKECLRYLRNNHRISVAASKLANMLG
YRGMGLSVGSMICWDKQLYYVDENGNRFSGQMFSGSGNSYAYAVDSGLREDTIVE
EAYDLGRGIVYATHRDAYSGGVYVMYHMMEDGWIKVCQDDVSQLIHLKEMF"
complement(7521..7609)
/gene="psmb8"
/number=6
complement(7610..7916)
/gene="psmb8"
/number=5
complement(7917..8112)
/gene="psmb8"
/number=5
complement(8113..8860)
/gene="psmb8"
/number=4
complement(8861..8990)
/gene="psmb8"
/number=4
complement(8991..9166)
/gene="psmb8"
/number=3
complement(9167..9278)
/gene="psmb8"
/number=3
complement(9279..9459)
/gene="psmb8"
/number=2
complement(9460..9601)
```

```
/gene="psmb8"
/number=2
complement(9602..9822)
/gene="psmb8"
/number=1
complement(9823..9972)
/gene="psmb8"
/number=1
complement(10598..12501)
/gene="psmb10"
complement(join(10598..10709,10825..10976,11108..11166,
11281..11396,11538..11678,11788..11885,12136..12223,
12440..12501))
/gene="psmb10"
complement(join(10598..10709,10825..10976,11108..11166,
11281..11396,11538..11678,11788..11885,12136..12223,
12440..12501))
/gene="psmb10"
/codon_start=1
/product="low molecular mass polypeptide subunit PSMB10"
/protein_id="CAC13118.1"
/db_xref="GI:10803371"
/db_xref="GOA:Q9DES1"
/db_xref="SPTREMBL:Q9DES1"
/translation="MALSNVLETAAGFNFDNAARNALRGLFEGGKPKPMKTGTTI
AGVFKDGVVLGADTRATSVSEVADKMKAKIHYPNIYCCGAGTAADTKTDLSS
NLTFISLNGSRKPRVMAVNILODTLYRHQIGANLILGGVDCGTHLYKVGPGYGV
DKVPYLTMGSDLAALGILLEDGFKHMEVERATELVRLAHAGIMSLGSGNNIDICV
ITRDVDYIRPYEESSEYDKSRTRYKYPGTPVLTKEVVPKLKLEMQETVQRMDTV"
complement(10598..10709)
/gene="psmb10"
/number=8
complement(10710..10824)
/gene="psmb10"
/number=7
complement(10825..10976)
/gene="psmb10"
/number=7
complement(10977..11107)
/gene="psmb10"
/number=6
complement(11108..11166)
/gene="psmb10"
/number=6
complement(11167..11280)
/gene="psmb10"
/number=5
complement(11281..11396)
/gene="psmb10"
/number=5
complement(11397..11537)
/gene="psmb10"
/number=4
complement(11538..11678)
/gene="psmb10"
/number=4
complement(11679..11787)
/gene="psmb10"
/number=3
complement(11788..11885)
```

```
Query Match      5.0%; Score 75.4; DB 5; Length 22398;
Best Local Similarity 45.0%; Pred. No. 0.0028;
Matches 329; Conservative 0; Mismatches 396; Indels 6; Gaps 1;

Qy      172 AATGCAATGTATCTGATTCGAATAAGATGAGGCTTCTTATTATTCCTCAAGTAAA 231
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      15368 AACGTCAATTTACATGATTTGGCGGATTACACGTTTAAACAGTACTTCTTCTGTTGT 15427

Qy      232 GAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTAATGGA 291
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      15428 GCATTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTT 15487
```



```
QY 292 GGGAGACATATGTAACTAATAAGATAGTGGTCAGCAATGAGATTGCGACATGGGCT 351
Db 15488 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 15547
QY 352 AAATCTATATCAGCTAACTACTACACAGTTTCCACTGTTTCTACTGAGTCAAAATATGATGT 411
Db 15548 GCTTCTACACATGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 15607
QY 412 ACTGAGGTATTAATGTTTCCAAATATGATATTAATGTTTCTAGCACTGTTTAATAAT 471
Db 15608 ACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 15667
QY 472 GGAGCTGTAATTAATGTTTCACTCTGTAATCTCAATGCTACTACTATTATCATGAAGAATACT 531
Db 15668 ACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 15727
QY 532 GATGCGACATGGGAGATGGTGGTGAATAAATCTGTAGATCAAAAAACGTTACTCGGTGCT 591
Db 15728 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCT 15781
QY 592 GATACAGTCAAAATATTAATTAAGAAATGAGTCAATTAATCATGGTACAGAAAAA 651
Db 15782 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 15841
QY 652 GGTATCAATATGTTATAAAGATACTATGCCATCTGTTCTGTAGTTGATTTGAACGAA 711
Db 15842 GCTTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 15901
QY 712 GGGTCTTATGAAGTAACTATTAATGATGATGAGGAAATATTACAACTTAACCTCAAGT 771
Db 15902 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 15961
QY 772 TCGGAAAAAGCACTGGGAAGTATAACCTGTTTAGAGGAAAAATAATTAATTTCAAGATTACT 831
Db 15962 TCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 16021
QY 832 ATTCGTTGGGAGCTACCAATFACTCAACCGGAATATCTCAAAATGGAGCTTAATGATGAC 891
Db 16022 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 16081
QY 892 TTTTATTATAA 902
Db 16082 AATAATATAA 16092
```

## RESULT 24

BX088589/c

LOCUS

DEFINITION BX088589 245292 bp DNA linear HTG 24-OCT-2003

Danio rerio clone DKEX-27E7, WORKING DRAFT SEQUENCE, 5 unordered

pieces.

ACCESSION BX088589

VERSION BX088589.5 GI:37988091

KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 245292)

Direct Submission

Submitted (23-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Oct 24, 2003 this sequence version replaced gi:37051062.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

----- Project Information

Center project name: zk27E7

----- Summary Statistics

Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 243531 bases at least Q40  
Consensus quality: 243896 bases at least Q30  
Consensus quality: 244209 bases at least Q20  
Insert size: 244892; sum-of-contigs  
Insert size: 237259; 8.1% error; agarose-fp  
Quality coverage: 7.91x in Q20 bases; sum-of-contigs Quality  
coverage: 8.16x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 21084: contig of 21084 bp in length  
\* 21085 21184: gap of 100 bp  
\* 21185 156057: contig of 134873 bp in length  
\* 156058 156157: gap of 100 bp  
\* 156158 197709: contig of 41552 bp in length  
\* 197710 197809: gap of 100 bp  
\* 197810 209759: contig of 11950 bp in length  
\* 209760 209859: gap of 100 bp  
\* 209860 245292: contig of 35433 bp in length.

## FEATURES

## source

1. 245292  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEX-27E7"  
/clone\_lib="DanioKey"

## misc\_feature

1. 21084  
/note="assembly\_fragment:02624"

## misc\_feature

fragment chain:1"  
21185..156057  
/note="assembly\_fragment:01829"

## misc\_feature

fragment chain:1"  
156158..197709  
/note="assembly\_fragment:01215"

## misc\_feature

fragment chain:1"  
197810..209759  
/note="assembly\_fragment:02518"

## misc\_feature

209860..245292  
/note="assembly\_fragment:00206.0"

## ORIGIN

Query Match 5.0%; Score 75.2; DB 2; Length 245292;  
Best Local Similarity 44.2%; Pred. No. 0.0014;  
Matches 408; Conservative 0; Mismatches 508; Indels 8; Gaps 2;  
QY 186 TGATTCGAATAAAGATGGAGCTTCTTATTTCCTCAAGGTAAGAAGCTGAGTATAA 245  
Db 8998 TAAATTATAATAATAATAATTTCTACTACTACTACTACTACTACTACTACTACTAA 8939  
QY 246 AGCTTCACTGATTTTAACTTCTTTTACGACAACTACTAATGGAGGAGACATATGT 305  
Db 8938 TAAATCTACTAATAATAATACTTCTACTGCTACTACTACTACTACTACTAATATCTTC 8879  
QY 306 AACTAAAAAGATACCTGGTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGC 365  
Db 8878 TACTACTAATAATACTTCTACTAATACTACTACTACTACTACTACTACTACTACTAA 8819  
QY 366 TAATACTACACAGTTTCCACTGTTTCTGAGTCAAAATAATGATGCTACTGAGGTTATTAA 425  
Db 8818 TGATACCTTCCATTAATACTACTACTACTACTACTACTACTACTACTATAATAATAA 8759  
QY 426 TGTTCCTCCCAATATGATATTATTTCTAGCACTGTTTAATATGAGCTGTTAATTAT 485  
Db 8758 CACCCCTTCTTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTAA 8699





\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

*	1	19051:	contig of 19051 bp in length
*	19052	19151:	gap of 100 bp
*	19152	21690:	contig of 2539 bp in length
*	21691	21790:	gap of 100 bp
*	21791	26578:	contig of 4788 bp in length
*	26579	26678:	gap of 100 bp
*	26679	45742:	contig of 13064 bp in length
*	45743	45842:	gap of 100 bp
*	45843	50124:	contig of 4282 bp in length
*	50125	50224:	gap of 100 bp
*	50225	78285:	contig of 28061 bp in length
*	78286	78385:	gap of 100 bp
*	78386	84233:	contig of 5954 bp in length
*	84240	84339:	gap of 100 bp
*	84340	180162:	contig of 95823 bp in length
*	180163	180262:	gap of 100 bp
*	180263	182961:	contig of 2699 bp in length
*	182962	183061:	gap of 100 bp
*	183062	193698:	contig of 16637 bp in length

FEATURES	Location/Qualifiers
103002	153004

```
1. 199698
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-222F23"
/clone_lib="CHORI-211"
```

## ORIGIN

Query Match 4.9%; Score 74; DB 2; Length 199698;  
Best Local Similarity 44.3%; Pred. No. 0.0024;  
Matches 302; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

137 ATAAAGCATATAAAGTTTTTGATGCAGAAATAGATAATGCAAAATGTATCTGATTCGAATA 196

[illegible]

107 A G C G C C C C C C C C C C T C A A G C T A A A G C T T C A A C T G 256

**QY**

\* NOTE: This is a 'working draft' sequence. It currently



/translation="MEEQNEEIEKTVSESEINEEVEEELFLPASSNDNLVKKESLRI  
ILSLLEKEFIDHIIKNNKKYKLTLLISSNRCCLLTLECLNLISAQPKELTERNLI  
EILFDMKEEIEIENDICDLYIMIIISNLTCKEGYKVLDDINNDNINIKEDNFKVSF  
FLNKLRYEFLPKISINKNLSKDYIYVSHVLINISIKESIVFKNVAFLNKISDQI  
LNVESRPAILPFIINLCINEVTHPIYHDDCVLPYVLSYLVTDYNTDKSGSYNNNS  
NNEINTQNIHIIWNKSSILVSCSVIKSRILILFYLGNRDSRNDIENVTKNLIEG  
SNSRGAIAAT"

repeat\_region

4677..4714

/rpt\_type=tandem

repeat\_region

5600..5641

/rpt\_type=tandem

repeat\_region

5745..5791

/rpt\_type=tandem

repeat\_region

6108..6152

/rpt\_type=tandem

repeat\_region

6168..6222

/rpt\_type=tandem

repeat\_region

6220..6268

/rpt\_type=tandem

repeat\_region

6420..6481

/rpt\_type=tandem

repeat\_region

6563..6621

/rpt\_type=tandem

repeat\_region

6718..6780

/rpt\_type=tandem

repeat\_region

6719..6790

/rpt\_type=tandem

repeat\_region

7543..7568

/rpt\_type=tandem

repeat\_region

7761..7816

/rpt\_type=tandem

repeat\_region

7824..7849

/rpt\_type=tandem

repeat\_region

7871..7930

/rpt\_type=tandem

gene

complement(<7953..>8438)

mRNA

locus tag="PF14\_0240"

CDS

complement(<7953..>8438)

complement(7953..8438)

locus tag="PF14\_0240"

locus tag="PF14\_0240"

codon\_start=1

product="ribosomal protein l21e, putative"

/protein\_id="AA036853.1"

/db\_xref="GI:23497308"

/translation="MGKSKGSGTGYKFKPRKHGECTANKYLEKLVGDYVDIV  
CDSTQQGMFPNYTHGKIGKLFHITKRGVGVNKRVRHIEQKKVCVRLEHVRKSR  
NEDFLLRKIKNAELIKEAKLNEHINIKRTGEPKPAAMIKVPSPKIIITIEPLPFEE  
Y"

repeat\_region

8467..8550

/rpt\_type=tandem

repeat\_region

8473..8539

/rpt\_type=tandem

repeat\_region

8477..8537

/rpt\_type=tandem

repeat\_region

8723..8844

/rpt\_type=tandem

repeat\_region

8723..8840

/rpt\_type=tandem

repeat\_region

8782..8844

/rpt\_type=tandem

repeat\_region

9019..9059

/rpt\_type=tandem

repeat\_region

9019..9058

/rpt\_type=tandem

repeat\_region

9133..9189

/rpt\_type=tandem

repeat\_region

9221..9254

/rpt\_type=tandem

repeat\_region

9604..9637

/rpt\_type=tandem

repeat\_region

9668..9723

/rpt\_type=tandem

repeat\_region

10036..10073

repeat\_region

10445..10472

repeat\_region

10511..10551

gene

complement(<10565..>11080)

mRNA

locus tag="PF14\_0241"

CDS

complement(10565..11080)

locus tag="PF14\_0241"

codon\_start=1

product="basic transcription factor 3b, putative"

/protein\_id="AA036854.1"

/db\_xref="GI:23497309"

/translation="MEKISPEILAAARAKLEKMGNNLROIGGKSGARRKIKKVKHNSI  
SNEKKINIILKKIGAGYGDVDEICVTRTGDTFLEPKPKLCLASLOSNTYIVTGKNE  
HKIDINKLFEGLKGNKNLDMNLLEKINIKNDPNIKNILNKESGDTPKREDEQEBANDVPD  
LVNFERVSKS"

repeat\_region

11429..11458

repeat\_region

11514..11584

repeat\_region

11514..11579

Query Match

Best Local Similarity

Matches 243; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

Score 73.8; DB 3; Length 250029;

4.9%; Pred. No. 0.0024;

156 TGATGAGAAATAGATAATGCAAAATGATCTGATTCGAAATAAGATGGAGCTTCTTATTT 215

48791 TGATAATAATAATGATAATAATAATAATGATAATGATAATAATAATAATAATAATCA 48850

216 AATTCCTCAAGGTAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTAC 275

48851 TAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 48910

276 GACAACTACTAATGGAGGGAGAACATATGTAACATAAAAGATATCGGTGAGCAATGA 335

48911 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATGATGA 48970

336 GATTCGGACATGGCTTAATCTATATCAGTAATAACTACACAGTTTCCACTGTTTACTGA 395

48971 TAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49030

396 GTCAATAATGATGCTACTGAGGTTTATTAATGTTTCCCAATATGGATATTTATGTTTC 455

49031 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49090

456 TAGCAGCTGTTAATAATGGAGCTGTAATTTATGTTTACATCTGTACTCCAAATGCTACTAT 515

49091 TGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49150

516 TCATGAAGAAGTACTGATCGACATGGGAGATGGTGGTGGAAAACTGTAGATCAAAA 575

49151 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 49210

576 AACGTACTCGTGGTGATACAGTCAAAATATATCTATTACTTTAAGAGTGCAGTCAATTA 635

49211 CAATAATAATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 680

636 TCATGTACAGAAAAGTGTATCAATATGTTTATAAGGATCTAT 680

49271 TAATAACAATAATAATGATAATAATAATAATAATAATAATAATAATAATGATAATGT 49315

RESULT 29

AF413050S2/c

LOCUS AF413050S2

DEFINITION Zea mays CAG and TAG long trinucleotide repeat.

ACCESSION AF413051

VERSION AF413051.1 GI:18001263

666 bp DNA linear

PLN 31-DEC-2001

KEYWORDS  
SEGMENT  
SOURCE  
ORGANISM

2 of 2  
Zea mays  
Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

1 (bases 1 to 666)  
Ananiev,E.V., Lorentzen,J. and Bruggemann,E.  
Microsatellite megatracts in the maize (Zea mays L.) genome  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

2 (bases 1 to 666)  
Ananiev,E.V., Lorentzen,J. and Bruggemann,E.  
Direct Submission  
Submitted (24-AUG-2001) Molecular Genetics, Pioneer Hi-Bred, 62-nd  
Avenue, Johnston, IA 50131, USA

FEATURES  
source

1..666  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4577"  
/note="present on chromosomes 1, 2 and 4"

misc\_feature  
1..666  
/note="3' end of a 2.6 kbp fragment"

repeat\_region  
1..537  
/note="CAG and TAG long trinucleotide repeat; organized in  
clusters"  
/rpt\_type=tandem  
/rpt\_unit="cag:tag"  
538..566  
/note="similar to LTR of Zeon 1 retrotransposon of Zea  
mays"

misc\_feature

ORIGIN

Query Match 4.8%; Score 73; DB 8; Length 666;  
Best Local Similarity 47.3%; Pred.No.0.021;  
Matches 220; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 100 ACAATTACAGTTCAAGATACCTCAAAAGCGCAACTATAAAGCATATAAAGTTTTGAT 159  
DB 540 ACAACTACTACTGGTACTACTACTACTACTACTACTACTACTACTACTACTACT 481  
QY 160 GCAGAAATAGATAAGCAATGATCTGATTCGATATAAGATGGAGCTTCATTAAATT 219  
DB 480 ACTGCTACTGCTACTACTACTACTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 421  
QY 220 CCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTACGACA 279  
DB 420 GCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 361  
QY 280 ACTACTAATGGAGGAGAACATATGTAATAAAGATACCTGCTGCTGCTGCTGCTGCT 339  
DB 360 ACTACTGCTGCTACTACTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTACT 301  
QY 340 GCGACATGGCTAAATCTATATCAGTAATACCTACACAGTTTCCACTGTTACTGAGTCA 399  
DB 300 GCTACTCTCTGCTACTGCTACTGCTACTGCTACTGCTACTGCTGCTGCTACTGCTACT 241  
QY 400 AATAATGATGCTACTGAGGTAAATGATTTTCCCAATATGGATATATTATGTTCTTAGC 459  
DB 240 GCTACTGCTGCTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTACTATTGCTACT 181  
QY 460 ACTGTTAATAATGGAGCTGTAATATGTTAGTTACATCTGTAACCTCCAAATGCTACTATTCA 519  
DB 180 GCTACTGCTACTGCTACTACTACTATTGCTACTGCTACTGCTACTGCTACTGCTACT 121  
QY 520 GAAAGAACTATGATCGGACATGGGAGATGGTGGTGGGAAACT 564  
DB 120 GCTACTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 76

LOCUS  
DEFINITION

199347 bp DNA linear HTG 24-SEP-2003  
Danio rerio clone DKEYP-29A10, WORKING DRAFT SEQUENCE, 12 unordered  
pieces.

ACCESSION  
VERSION

BX321884  
BX321884.5 GI:35209510

KEYWORDS  
SOURCE  
ORGANISM

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
Danio rerio (zebrafish)  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
1 (bases 1 to 199347)  
McLaren,S.  
Direct Submission  
Submitted (23-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Sep 24, 2003 this sequence version replaced gi:3260915.  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zkp29A10  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 195192 bases at least Q40  
Consensus quality: 196456 bases at least Q30  
Consensus quality: 197320 bases at least Q20  
Insert size: 198247; sum-of-contigs  
Insert size: 172014; 5.6% error; agarose-fp  
Quality coverage: 7.64x in Q20 bases; sum-of-contigs Quality  
coverage: 10.37x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 94733: contig of 94733 bp in length  
\* 94734 94833: gap of 100 bp  
\* 94834 139839: contig of 45006 bp in length  
\* 139840 139939: gap of 100 bp  
\* 139940 142547: contig of 2608 bp in length  
\* 142548 142647: gap of 100 bp  
\* 142648 145652: contig of 3005 bp in length  
\* 145653 145752: gap of 100 bp  
\* 145753 147869: contig of 2117 bp in length  
\* 147870 147969: gap of 100 bp  
\* 147970 150959: contig of 2990 bp in length  
\* 150960 151059: gap of 100 bp  
\* 151060 154357: contig of 3298 bp in length  
\* 154358 154457: gap of 100 bp  
\* 154458 156995: contig of 2538 bp in length  
\* 156996 157095: gap of 100 bp  
\* 157096 159488: contig of 2393 bp in length  
\* 159489 159588: gap of 100 bp  
\* 159589 161652: contig of 2064 bp in length  
\* 161653 161752: gap of 100 bp  
\* 161753 165174: contig of 3422 bp in length  
\* 165175 165274: gap of 100 bp  
\* 165275 199347: contig of 34073 bp in length.  
Location/Qualifiers  
1..199347  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEYP-29A10"

FEATURES  
source

1..199347  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEYP-29A10"

misc_feature		/clone lib="DanieKeypilot"
misc_feature	1. .94733	/note="assembly_fragment:00836
	fragment_chain:1	94834.139839
misc_feature	/note="assembly_fragment:00061	fragment_chain:1
	139940.142547	/note="assembly_fragment:00422"
misc_feature	142648.145652	/note="assembly_fragment:00804"
	145753.147869	/note="assembly_fragment:00868"
misc_feature	147970.150959	/note="assembly_fragment:01524"
	151060.154357	/note="assembly_fragment:01995"
misc_feature	154458.156995	/note="assembly_fragment:02125"
	157096.159488	/note="assembly_fragment:02132"
misc_feature	159589.161652	/note="assembly_fragment:02163"
	161753.165174	/note="assembly_fragment:02166"
misc_feature	163275.199347	/note="assembly_fragment:02257.0"

ORIGIN

Query Match 4.8%; Score 73; DB 2; Length 199347;  
Best Local Similarity 43.9%; Pred. No. 0.0035;  
Matches 446; Conservative 0; Mismatches 565; Indels 4; Gaps 3;

Qy	110	TTCAAGTACTCAAAAAGGCGCACCTATAAGCATATAAGCTTTTGTGACGAGAAATAG	169
Db	168472	TTGTATCTAGTTTCAAAATGAAATTTATTACTCAGACATATTGTCTTTTAATAATAATA	168531
Qy	170	ATATGCAATGTATCTGATTTCGAATAAGATGAGCTTCTTATTAAATTCCTCAAGGTA	229
Db	168532	ATAATAATAATAATAATAATAGTAAATTAATAGTAAATAATAATAATAATAATAATA	168591
Qy	230	AAGAAGCTGAGTATAAAGCTTCAACTGATTTAAATCTCTTTTACGACAACTACTAATG	289
Db	168592	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	168651
Qy	290	GAGGAGACATATGTACTAATAAAGATCTGCGTCAGCAATGAGATTGCGCATGGG	349
Db	168652	ATAATAGTAGTAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	168711
Qy	350	CTAAATCTATACAGCTAATACTACACCAGTTTCCACTGTCTACTGAGTCAATAATGATG	409
Db	168712	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	168771
Qy	410	GTACTGAGGTTATTAATGTTTCCCAATAGATGATTTATTATGTTCTTAGCAGCTTTAATA	469
Db	168772	ATAATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	168831
Qy	470	ATGGAGCTGTAATATTGTTTACATCTGAACTCCAAATGCTACTATTCTATGAAAGAAATA	529
Db	168832	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	168891
Qy	530	CTGATCGCAGATGGGAGATGGTGGGAAACCTGTAGATCAAAAACGCTACTCGGTTG	589
Db	168892	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	168951
Qy	590	GTGATACAGTCAATATATCTACTTATAGATGCGAGTCAATTTATCATGGTACAGAAA	649
Db	168952	TAAATAAAGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	169011
Qy	650	AAAGTGATCAATATGTTTAAAGGATCTATGCCATCTGCTTCTGTAGTTGATTTGACG	709
Db	169012	GTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	169071

Search completed: July 27, 2004, 00:04:31  
Job time : 6040 secs

Qy	710	AAGGCTCTTATGAAGTAACCTATTACTGATGGATCAGGGAATATTACAACCTCAACTCAAG	769
Db	169072	ATAGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	169131
Qy	770	GTTCCGAAAAAGCAACTGGGAAGTATAACCTGTTAGAGGNA--AATAATAATTTCCAGATT	828
Db	169132	ATAATAATAAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	169191
Qy	829	ACTATTCGGTGGGCGAGCTACCAATACTCCAAACCGGAAATACTCAAATGGAGCTAATGAT	888
Db	169192	AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	169251
Qy	889	GACTTTTTTTTATAGGGAATAATAATACAAATCAACAGTCACTTATACAGGAGTATTAAAGAGT	948
Db	169252	AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	169309
Qy	949	GGAGCTAAACAGGTTTCAGCTGATTTTACCAGAAATAACAACATTCGGACCACCAACCCC	1008
Db	169310	AGTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	169369
Qy	1009	AATCTAGCAATGATGACCCAGGTCAAAAAGTAAACAGTGGGGATGGTCAAAATTTACTATA	1068
Db	169370	ATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	169428
Qy	1069	AAAAAAATTTGATGGTTCCACAAAAGCTTTCATTACAAGTGTCTATATTGTTTTAA	1123
Db	169429	TTAAAAATAAAAAATAACAACAACAATAATAATAATAATAATAATAATAATAATAATA	169483



GenCore version 5.1.6  
 -Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 00:07:57 ; Search time 666 Seconds  
 (without alignments)  
 9625.422 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaataatgattca.....gtcgtcgtttacgtttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10\_0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : N\_Geneseq\_29Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001as:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002as:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1509	100.0	1509	5	AAD02390	Aad02390 Virulent
2	1509	100.0	1509	6	ABK11585	Abk11585 S. agalac
3	64.8	4.3	1983	2	AAX20108	Aax20108 Enterococ
4	64.8	4.3	1983	6	ABN98093	Abn98093 E. faecali
5	64.8	4.3	1983	7	ACA88057	Abx61663 Enterococ
6	64.8	4.3	1983	7	ABX61663	Abx61663 Enterococ
7	64.8	4.3	1983	7	ABK11591	Abk11591 E. faecali
8	64.8	4.3	15614	2	AAX12982	Aax12982 Enterococ
9	64.8	4.3	15614	6	ABN98077	Abn98077 Enterococ
10	59.6	3.9	11922	3	ACA87018	Aca87018 Plasmodi
11	59.2	3.9	1137	7	ACA28118	Abn98777 Plasmodi
12	56.6	3.8	969	6	ABQ39490	Abq39490 Oligonuc
13	56.6	3.8	969	6	ABQ39491	Abq39491 Oligonuc
14	55.4	3.7	2408	4	ABL28164	Abi28164 Drosophi
15	55.2	3.7	32392	6	ABL56203	Abi56203 AmEVP gen
16	54	3.6	408	4	ABL28165	Abi28165 Drosophi
17	53.6	3.6	2535	4	ABL25476	Abi25476 Drosophi
18	53	3.5	4985	6	ABQ75107	Abq75107 Anopheles
19	53	3.5	4985	9	ACF79720	Abn67916 Streptoco
20	53	3.5	6033	3	AAA70152	Aaa70152 Plasmodi
21	51.2	3.4	2703	6	ABN67916	Abn67916 Streptoco
22	51.2	3.4	110000	6	ABN71527_13	Continuation (14 o
23	50.4	3.3	14066	2	AAX99556	Aax99556 Nucleic a

RESULT 1

## ALIGNMENTS

24	50.4	3.3	14067	7	ACA52811	AcA52811 Prokaryot
25	50.2	3.3	7143	3	AAA70250	Aaa70250 Plasmodi
26	50	3.3	3996	6	AAD47004	Aad47004 Plasmodi
27	49.8	3.3	1687	6	AAX20109	Aax20109 Enterococ
28	49.8	3.3	1687	6	ABN98094	Abn98094 E. faecali
29	49.8	3.3	1687	7	ACA88058	Abn98058 E. faecali
30	49.8	3.3	1687	7	ABX61664	Abx61664 Enterococ
31	49.6	3.3	3837	3	AAA70211	Aaa70211 Plasmodi
32	48.6	3.2	4997	7	ACA47636	AcA47636 Prokaryot
33	48.6	3.2	4851	6	AAS20800	Aas20800 Clostridi
34	48.4	3.2	4197	2	AAQ99430	Aaq99430 B. sphaer
35	47.8	3.2	740	4	AAAC85918	Aac85918 rCP41 cDN
36	47.8	3.2	1083	5	AAS76745	Aas76745 DNA encod
37	47.8	3.2	8244	5	AAS66529	Aas66529 DNA encod
38	47.6	3.2	11143	4	ABL12834	Abi12834 Drosophi
39	47.4	3.1	1037	3	AAA59242	Aaa59242 Exons D.
40	47.4	3.1	1472	3	AAA59241	Aaa59241 Exons D.
41	47.4	3.1	2428	7	ADA89818	Ada89818 Staphyloc
42	47.4	3.1	3945	2	AAAX77593	Aax77593 S. aureus
43	47.4	3.1	4047	4	AAS52261	Aas52261 Staphyloc
44	47.4	3.1	4047	7	ACF73996	AcF73996 Staphyloc
45	47.4	3.1	4050	4	AAS55402	Aas55402 Staphyloc
46	47.4	3.1	4143	7	ABT14918	Abt14918 Pathogen
47	47.4	3.1	4418	2	AAV74765	Aav74765 Staphyloc
48	47.4	3.1	7458	3	AAA70106	Aaa70106 Plasmodi
49	47.2	3.1	5314	9	ADB46161	AdB46161 Plasmodi
50	47	3.1	3567	3	AAA70117	Aaa70117 Plasmodi
51	47	3.1	5940	3	AAA70105	Aaa70105 Plasmodi
52	46.8	3.1	1938	4	AAS53795	Aas53795 S. epider
53	46.8	3.1	1948	2	AAQ12693	Aaq12693 Toxin A g
54	46.8	3.1	2801	7	ACA46974	AcA46974 Prokaryot
55	46.8	3.1	3188	4	AAH54473	Aah54473 S. epider
56	46.8	3.1	3594	4	AAH54439	Aah54439 S. epider
57	46.8	3.1	3641	4	AAH54487	Aah54487 S. epider
58	46.8	3.1	4506	4	AAH53728	Aah53728 S. epider
59	46.8	3.1	4590	6	ABN91645	Abn91645 Staphyloc
60	46.8	3.1	7047	6	ABK28386	Abk28386 DNA trans
61	46.2	3.1	535	6	ABQ42292	Abq42292 Oligonuc
62	46.2	3.1	535	6	ABQ42293	Abq42293 Oligonuc
63	46.2	3.1	4158	7	ACA20270	AcA20270 Prokaryot
64	46	3.0	2394	9	ADC89717	Adc89717 L. johnso
65	46	3.0	2900	9	ADC89716	Adc89716 L. johnso
66	46	3.0	5139	3	AAA70139	Aaa70139 Plasmodi
67	45.8	3.0	1159	3	AAA59240	Aaa59240 An EcoRI
68	45.8	3.0	110000	6	ABQ67196_1	Continuation (2 of
69	45.8	3.0	110000	6	ABQ69245_26	Continuation (27 o
70	45.6	3.0	2017	9	ADB46163	AdB46163 Plasmodi
71	45.6	3.0	4423	2	AAQ29470	Aaq29470 Extracell
72	45.6	3.0	6397	4	AAS45400	Aas45400 Chemical
73	45.6	3.0	6397	4	AAS46459	Aas46459 Tumour su
74	45.6	3.0	6397	6	ABL33346	Abi33346 Human imm
75	45.6	3.0	6397	6	ABK28245	Abk28245 DNA trans
76	45.6	3.0	6744	2	AAQ29471	Aaq29471 Extracell
77	45.6	3.0	50000	6	ABL56202	Abi56202 AmEVP gen
78	45.4	3.0	2217	7	ACA29569	AcA29569 Prokaryot
79	45.4	3.0	2505	6	ABN66950	Abn66950 Streptoco
80	45.4	3.0	3927	3	AAA70101	Aaa70101 Plasmodi
81	45.4	3.0	110000	6	ABN71527_19	Continuation (20 o
82	45.2	3.0	861	6	ABL55632	Abi55632 AmEVP tri
83	45.2	3.0	3931	8	ACL84661	Acc84661 P. falcip
84	45.2	3.0	50000	6	ABL56201	Abi56201 AmEVP gen
85	45.2	3.0	110000	6	ABA90521_14	Continuation (15 o
86	45	3.0	12237	6	ABL34358	Abi34358 Human imm
87	44.8	3.0	16287	6	ABL32672	Abi32672 Human imm
88	44.6	3.0	954	6	ABQ76556	Abq76556 C. albica
89	44.6	3.0	11679	8	ADA30090	Ada30090 DNA encod
90	44.4	2.9	1998	3	AAA70212	Aaa70212 Plasmodi

```
AAD02390
ID AAD02390 standard; DNA; 1509 BP.
XX
AC AAD02390;
XX
DT 24-APR-2001 (first entry)
DE
DE Virulent group B Streptococcus agalactiae spb1 DNA.
XX
XX Type III virulent group B; spb1; cell wall bound protein; antibacterial;
KW immunisation; group B streptococci; GBS infection; vaccine; bacteraemia;
KW pneumonia; meningitis; endocarditis; osteoarticular infection; ds.
XX
OS Streptococcus agalactiae.
XX
PH Key Location/Qualifiers
FT CDS 1..1509
FT /*tag= a
FT /product= "S. agalactiae spb1 protein"
FT sig_peptide 19..87
FT /*tag= b
FT mat_peptide 88..1506
FT /*tag= c
FT /product= "S. agalactiae mature spb1 protein"
XX
XX WO2000787-A1.
XX
XX 28-DEC-2000.
XX
XX 21-JUN-2000; 2000WO-US017082.
XX
XX 21-JUN-1999; 99US-0140084P.
XX
XX (UTAH ) UNIV UTAH RES FOUND.
XX
XX Adderson E, Bohnsack J;
XX
XX WPI; 2001-102693/11.
XX P-PSDB; AAY72357.
XX
XX Polynucleotide from spb1 and 2 genes derived from virulent Group B
XX streptococci, polypeptide encoded by the polynucleotide useful as vaccine
XX for immunizing a mammal against the streptococcal infection.
XX
XX Claim 9; Page 20-23; 34pp; English.
XX
XX The present sequence is spb1 gene from type III virulent group B
XX Streptococcus agalactiae. The spb1 protein has the characteristics of a
XX cell wall bound protein and has antibacterial activity. The N-terminus
XX of the spb1 protein is a hydrophilic, basic stretch of 6 amino acids
XX followed by a 23 amino acid hydrophobic, proline rich core, consistent
XX with a signal peptide. The hydrophilic mature protein terminates in
XX atypical LPXTG domain that immediately precedes a hydro- phobic 20 amino
XX acid core and a short, basic hydrophilic terminus. The spb1 protein is
XX used as a vaccine to immunise mammals against group B streptococci (GBS)
XX infection (e.g. bacteraemia, pneumonia, meningitis, endocarditis and
XX osteoarticular infections). Determination of the gene products specific
XX to type III-3 GBS is useful for diagnosing mammals infected or colonised
XX by virulent GBS
XX
XX Sequence 1509 BP; 534 A; 241 C; 298 G; 436 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 1509; DB 5; Length 1509;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTTACATTTGGTATGGCTGTA 60
XX |||||||
XX 1 ATGAAAAGAAATGATTCAATCGCTGTAGTCGCGAGTTTACATTTGGTATGGCTGTA 60
XX |||||||
XX 61 TCACCAAGTACGCGGATAGCTTTTGGCGCTGAGACAGGACAAATTACAGTTCAAGATAC 120
XX |||||||
XX 61 TCACCAAGTACGCGGATAGCTTTTGGCGCTGAGACAGGACAAATTACAGTTCAAGATAC 120
XX |||||||

121 CAAAAGCGCAACCTATAAGCATATAAAGTTTTTGTGTCAGAAATAGATAATGCAAAAT 180
121 CAAAAGCGCAACCTATAAAGCATATAAAGTTTTTGTGTCAGAAATAGATAATGCAAAAT 180
181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTAAATTCCTCAAGGTAAGAAGCTGAG 240
181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTAAATTCCTCAAGGTAAGAAGCTGAG 240
241 TATAAAGCTTCAACTGATTTTAAATTTCTTTTACGCAACTACTAATGAGGGGGAACA 300
241 TATAAAGCTTCAACTGATTTTAAATTTCTTTTACGCAACTACTAATGAGGGGGAACA 300
301 TATGTAACATAAAAAGATACATCGCGTCAGCAAAATGAGATTGGCAGATGGGCTAAATCTATA 360
301 TATGTAACATAAAAAGATACATCGCGTCAGCAAAATGAGATTGGCAGATGGGCTAAATCTATA 360
361 TCAGCTAATACACACAGTTTCCACTGTGTTCTAGCAGCTGTTTAATAATGGAGCTGTA 480
361 TCAGCTAATACACACAGTTTCCACTGTGTTCTAGCAGCTGTTTAATAATGGAGCTGTA 480
421 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCAGCTGTTTAATAATGGAGCTGTA 480
421 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCAGCTGTTTAATAATGGAGCTGTA 480
481 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCAGCTGTTTAATAATGGAGCTGTA 540
481 ATTAATGTTTCCCAATATGGATATTTATGTTTCTAGCAGCTGTTTAATAATGGAGCTGTA 540
541 TGGGAGATGGTGGGAAACCTGTAGATCAAAAACGTACTCGGTTGGTGATACAGTC 600
541 TGGGAGATGGTGGGAAACCTGTAGATCAAAAACGTACTCGGTTGGTGATACAGTC 600
601 AAATATACCTATTACTTATAAGAATGCAGTCAATATCATGTTACGAGGAGGAGTATCAA 660
601 AAATATACCTATTACTTATAAGAATGCAGTCAATATCATGTTACGAGGAGGAGTATCAA 660
661 TATGTTATAAGGATACATATGCCATCTGCTCTGTAGTTGATTTCGAAAGGAGGCTTAT 720
661 TATGTTATAAGGATACATATGCCATCTGCTCTGTAGTTGATTTCGAAAGGAGGCTTAT 720
721 GAAAGTAACTATTACTGTAGTGGATCAGGGAATATTCAAACTCAAGGTTTCGGAATA 780
721 GAAAGTAACTATTACTGTAGTGGATCAGGGAATATTCAAACTCAAGGTTTCGGAATA 780
781 GCAACTGGGAAGTATAAAGTGTAGAGGAAATAATAATTCACGATTACTATTCGCTGG 840
781 GCAACTGGGAAGTATAAAGTGTAGAGGAAATAATAATTCACGATTACTATTCGCTGG 840
841 GCAGCTACCAATACCTCCAAACCGGAATCTCAAAATGGAGCTAATGATGACATTTTAT 900
841 GCAGCTACCAATACCTCCAAACCGGAATCTCAAAATGGAGCTAATGATGACATTTTAT 900
901 AAGGGAATAAATAACATCACAGTCACTTATACAGGAGTATTAAAGAGTGGAGCTAAACCA 960
901 AAGGGAATAAATAACATCACAGTCACTTATACAGGAGTATTAAAGAGTGGAGCTAAACCA 960
961 GGTTCAGCTGATTTACCGAATAAATAACATCACAGTGGAGTCAATCAACCCCAATAGTCAAT 1020
961 GGTTCAGCTGATTTACCGAATAAATAACATCACAGTGGAGTCAATCAACCCCAATAGTCAAT 1020
1021 GATGACCCAGGTCAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1080
1021 GATGACCCAGGTCAAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 1080
1081 GGTTCACAAAAGCTTCATTACAGAGTCTATATTTGTTTTTAAAGAGTCTACGGGTCAA 1140
1081 GGTTCACAAAAGCTTCATTACAGAGTCTATATTTGTTTTTAAAGAGTCTACGGGTCAA 1140
1141 TTTCTAAACTTTAACGATAAAATACGTTGAATGGGCGACAGAGCTTAATGCAACAGAA 1200
1141 TTTCTAAACTTTAACGATAAAATACGTTGAATGGGCGACAGAGCTTAATGCAACAGAA 1200
```

QY 1201 TATACACAGGAGCAGATGGTATTAATTACCATTTACAGGCTTGAAGAAGGTACATCTAT 1260  
DB 1201 TATACACAGGAGCAGATGGTATTAATTACCATTTACAGGCTTGAAGAAGGTACATCTAT 1260  
QY 1261 CTAGTTGAGAAAAGGCTCCCTTAGCTTACAAATTTGTTAGATAACTCTCAGAAGGTTATT 1320  
DB 1261 CTAGTTGAGAAAAGGCTCCCTTAGCTTACAAATTTGTTAGATAACTCTCAGAAGGTTATT 1320  
QY 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTATAGTTAAACCCAACTGTT 1380  
DB 1321 TTAGGAGATGGAGCCACTGATACGACTAATTCAGATAACCTTTTATAGTTAAACCCAACTGTT 1380  
QY 1381 GAAATATACAAAGGTACTGAGTTGCTTCAACAGGTGGTATGTTAGTACAACTTTTCTAC 1440  
DB 1381 GAAATATACAAAGGTACTGAGTTGCTTCAACAGGTGGTATGTTAGTACAACTTTTCTAC 1440  
QY 1441 ATTATAGGTGCAATTTTATAGTAATAGCAGCAGGTATCGTGTGTTGCTCGTCTGTTTA 1500  
DB 1441 ATTATAGGTGCAATTTTATAGTAATAGCAGCAGGTATCGTGTGTTGCTCGTCTGTTTA 1500  
QY 1501 CGTCTCTTAA 1509  
DB 1501 CGTCTCTTAA 1509  
RESULT 2  
ABK11585  
ID ABK11585 standard; DNA; 1509 BP.  
AC ABK11585;  
XX  
DT 05-JUN-2002 (first entry)  
DE S. agalactiae Sbp1 gene.  
XX  
KW Extracellular matrix adhesion; Ema; ds; group B streptococcus; GBS;  
KW DNA vaccine; Sbp1; Sbp2; Rib; Lmb; C5a-ase; C protein alpha antigen;  
KW neonatal bacterial infection; gene.  
XX  
OS Streptococcus agalactiae.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..1509  
FT /\*tag= a  
FT /\*product= "Sbp1"  
XX  
PN WO200212294-A2.  
XX  
PD 14-FEB-2002.  
XX  
PF 08-AUG-2001; 2001WO-US024795.  
XX  
PR 08-AUG-2000; 2000US-00634341.  
XX  
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
PA (UTAH-) UNIV UTAH RES FOUND.  
XX  
PI Adderson E. Bohnsack J;  
XX  
DR WPI; 2002-257465/30.  
DR P-PSDB; AAU77626.  
XX  
PT New streptococcal matrix adhesion (Ema) polypeptides, useful as vaccines,  
PT particularly for treating or preventing infections by virulent forms of  
PT streptococci.  
XX  
PS Example 1; Page 147-150; 177pp; English.  
XX  
CC The invention relates to isolated streptococcal polypeptides, which  
CC comprise the Streptococcus agalactiae extracellular matrix adhesion (Ema)  
CC polypeptides EmaA, EmaB, EmaC, EmaD or EmaE, and their encoding nucleic  
CC acids. Also included are the a vaccine comprising an Ema protein, an anti  
CC -Ema antibody, a pharmaceutical composition comprising one or more Ema

CC proteins plus an anti Ema antibody, optionally in combination with at  
CC least one antibody to a protein selected from Sbp1 and Sbp2, Rib, Lmb,  
CC C5a-ase or C protein alpha antigen, an immortal cell line producing an  
CC anti-Ema monoclonal antibody, a vector comprising the Ema nucleic acid, a  
CC host cell transformed with the vector, a nucleic acid vaccine comprising  
CC the vector and Ema homologous proteins and their encoding nucleic acids,  
CC from additional bacterial species (S. pneumoniae, S. pyogenes,  
CC Enterococcus faecalis and Corynebacterium diphtheriae). The streptococcal  
CC polypeptides are useful as vaccines, particularly for treating or  
CC preventing infections by virulent forms of streptococci, especially group  
CC B streptococci (GBS) the most common cause of serious bacterial disease  
CC in neonates. The present sequence is the S. agalactiae Sbp1 gene  
XX  
SQ Sequence 1509 BP; 534 A; 241 C; 298 G; 436 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1509; DB 6; Length 1509;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGAGTTTAGCATTTGGTATGGCTGTA 60  
DB 1 ATGAAAAGAAAATGATTCAATCGCTGTAGTGGAGTTTAGCATTTGGTATGGCTGTA 60  
QY 61 TCACCAAGTTACCGCGATAGCTTTTCCGCTGAGACAGGCAATACAGATTCAAGATACT 120  
DB 61 TCACCAAGTTACCGCGATAGCTTTTCCGCTGAGACAGGCAATACAGATTCAAGATACT 120  
QY 121 CAAAAGGCGCAACTATAAGCATATAAAGTTTTTGTATGAGAAAATAGATAATGCAAAAT 180  
DB 121 CAAAAGGCGCAACTATAAGCATATAAAGTTTTTGTATGAGAAAATAGATAATGCAAAAT 180  
QY 181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTAAGAAGCTGAG 240  
DB 181 GTATCTGATTCGAATAAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTAAGAAGCTGAG 240  
QY 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTAATAGGAGGAGACA 300  
DB 241 TATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTAATAGGAGGAGACA 300  
QY 301 TATGTAACATAAAGAGATACCTGCGTCAGCAATGAGATGGCATGGGCTAAATCTATA 360  
DB 301 TATGTAACATAAAGAGATACCTGCGTCAGCAATGAGATGGCATGGGCTAAATCTATA 360  
QY 361 TCAGCTAATACTACACCAAGTTTCCACTGTTTACTGAGTCAAAATAATGATGGTACTGAGTT 420  
DB 361 TCAGCTAATACTACACCAAGTTTCCACTGTTTACTGAGTCAAAATAATGATGGTACTGAGTT 420  
QY 421 ATTAATGTTTCCCAATATGATATATATGTTTCTAGCACTGTTTAAATAGGAGCTGTA 480  
DB 421 ATTAATGTTTCCCAATATGATATATATGTTTCTAGCACTGTTTAAATAGGAGCTGTA 480  
QY 481 ATTATGTTTACATCTGTAATCTCCAAATGCTACTATTTCATGAAAAGAAATACCTGATCGGACA 540  
DB 481 ATTATGTTTACATCTGTAATCTCCAAATGCTACTATTTCATGAAAAGAAATACCTGATCGGACA 540  
QY 541 TGGGAGATGGTGGTGGAAAACCTGTAGATCAAAAACGTTACTCGTGGTGGTATACAGTC 600  
DB 541 TGGGAGATGGTGGTGGAAAACCTGTAGATCAAAAACGTTACTCGTGGTGGTATACAGTC 600  
QY 601 AAATATACTATTACTTATAAGAAATGCAGTCAATTTATCATGTGACAGAAAAGTGTATCAA 660  
DB 601 AAATATACTATTACTTATAAGAAATGCAGTCAATTTATCATGTGACAGAAAAGTGTATCAA 660  
QY 661 TATGTTATAAAGGATACTATGCGCATCTGCTTCTGTAGTTGATTTGAACGAAGGGTCTTAT 720  
DB 661 TATGTTATAAAGGATACTATGCGCATCTGCTTCTGTAGTTGATTTGAACGAAGGGTCTTAT 720  
QY 721 GAAGTAACTATTACTGATGGATCAGGGAATATTCAACTCTAAGCTTCAGGTTTCGAAAAA 780  
DB 721 GAAGTAACTATTACTGATGGATCAGGGAATATTCAACTCTAAGCTTCAGGTTTCGAAAAA 780  
QY 781 GCAACTGGGAAGTATACTCTTTAGAGGAAAATTAATTTTCAGATTTACTATTCCGTGG 840  
DB 781 GCAACTGGGAAGTATACTCTTTAGAGGAAAATTAATTTTCAGATTTACTATTCCGTGG 840



```
XX 18-APR-2002.
XX 04-MAY-1996; 98US-00071035.
XX 04-MAY-1998; 98US-00071035.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX WPI: 2002-425450/45.
XX P-PSDB; ABP43337.
XX
XX New genes and polypeptides from Enterococcus faecalis, useful as vaccines
XX for preventing, treating or attenuating an infection caused by a member
XX of the Enterococcus genus in an animal, particularly E. faecalis.
XX Claim 1; Page 92; 255pp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of polypeptides from Enterococcus faecalis. The proteins can be
XX used as vaccines for preventing or attenuating an infection caused by a
XX member of the Enterococcus genus in an animal, particularly E. faecalis.
XX The polynucleotide is also useful for preventing or treating E. faecalis
XX infection. The present sequence is a coding sequence of the invention
XX
XX Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;
XX
XX Query Match 4.3%; Score 64.8; DB 6; Length 1983;
XX Best Local Similarity 52.1%; Pred. NO. 0.00013;
XX Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;
XX
QY 1180 ACAGAAGCTAATGCAACAGAAATATACACAGAGCAGATGGTATTAATACCATTTACAGGC 1239
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1240 TTGAAGAAGGTACATATCTAGTTGAGAAAAAGGCTCCCTTAGGTTTACAATTTGTTA 1299
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1720 CTTAAATACGGTACCTATTATTAGAGAAACTGTAGCTCCTGATGATTATGCTTTGTTA 1779
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1300 GATAACTCTCAGAGGTTATTTTAGGAGTGGAGCCACTGATACGACTAAATTCAGATAAC 1359
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1780 ACAATTCGGATTGAATTTGTGTCAATGAACAAATCATATGGCACAA-----CAGAAAC 1833
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1360 CTTTGTAGTTAACCCAACTGTTGAAATAACAAAGGTACTGAGTGTGCTTCAACAGGTGTT 1419
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1834 CTAGTTTACACAGAAAGTACCAACAAACAAAGGTACTTACCTTCAACAGGTGGC 1893
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1420 ATTGTACAACAATTTTCTACATTATAGTGCAATTTTAGTAAATAGGAGCAGGTATCGTG 1479
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1894 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTGCGAGGAGTCTAC 1953
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1480 CTTGTGCTCGTCGTTTACGTTCTTAA 1509
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1954 TTTGCTAGACGTAGAAAAGAAAATGCTTAA 1983
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
ACAB8057
ID ACAB8057 standard; DNA; 1983 BP.
XX
AC ACAB8057;
XX
DT 07-JUL-2003 (first entry)
XX
DE E. faecalis novel gene #109.
XX
XX Gene; db; endocarditis; bacteraemia; urinary tract infection; UTI;
XX intraabdominal infection; soft tissue infection; neonatal sepsis;
XX vaccine.
XX
OS Enterococcus faecalis.
```

Db 1780 ACAATCGGATTAATTTGGTCAATGAACAAATCATATGGCAAA-----CAGAAAC 1833  
QY 1360 CTTTATAGTTAAACCAACTGTTGAAATACAAAGGTACTGAGTTGCCCTTCAACAGGTGGT 1419  
Db 1834 CTAGTTTCACAGAAAGTACCAACAAACAAAGGTACCTTACCTTCAACAGGTGGC 1893  
QY 1420 ATTGGTACAAATTTTCTACATATATAGTGCAATTTTAGTAAATAGGAGCAGGTATCGTG 1479  
Db 1894 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTATGAGGAGTCTAC 1953  
QY 1480 CTTGTTGCTCGTCTGTTTACGTTCTTAA 1509  
Db 1954 TTTGCTAGACGTAGAAAAGAAATGCTTAA 1983  
RESULT 6  
ABX61663  
ID ABX61663 standard; DNA; 1983 BP.  
XX AC ABX61663;  
XX DT 26-FEB-2003 (first entry)  
XX DE Enterococcus faecalis EF040 polynucleotide #109.  
XX EF040; gene; ds; immunostimulant; antibacterial; gene mapping.  
XX OS Enterococcus faecalis.  
XX PN US6448043-B1.  
XX PD 10-SEP-2002.  
XX PF 04-MAY-1998; 98US-00071035.  
XX PR 06-MAY-1997; 97US-0044031P.  
XX PR 16-MAY-1997; 97US-0046655P.  
XX PR 14-NOV-1997; 97US-0066009P.  
XX PR 14-NOV-1997; 97US-0066009P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;  
XX WPI; 2003-089120/08.  
XX P-PSDB; ABU13616.  
XX New EF040 polypeptides and polynucleotides from Enterococcus faecalis,  
XX useful for generating an immune response against E. faecalis and other  
XX Enterococcus species, and as vaccines against other bacterial genera.  
XX Example 1; Col 113-116; 146pp; English.  
XX The invention relates to polynucleotide fragments of a gene from  
XX Enterococcus faecalis, EF040, and the polypeptides encoded by them. The  
XX polypeptides are useful in detecting E. faecalis, as epitope tags, as  
XX molecular weight markers on SDS-PAGE gels or for molecular sieve gel  
XX filtration columns, in generating antibodies that specifically bind to  
XX the E. faecalis polypeptides, in generating an immune response against E.  
XX faecalis and other Enterococcus species and as vaccines against other  
XX bacterial genera. The polynucleotides are useful as probes for gene  
XX mapping and for identifying E. faecalis in biological samples. Sequences  
XX CC ABX61555-ABX61802 represent EF040 polynucleotides of the invention. Note:  
XX The sequence data for this patent can also be obtained from USPTO at  
XX sequata.uspto.gov/sequence.html  
XX Sequence 1983 BP; 729 A; 330 C; 426 G; 498 T; 0 U; 0 Other;  
XX Query Match 4.3%; Score 64.8; DB 7; Length 1983;  
XX Best Local Similarity 52.1%; Pred. No. 0.00013;  
XX Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAAGCTAATGCAACAGAAATATACAACAGGAGCAGATGGTATATAATCCATTACAGGC 1239  
Db 1660 ACAAGCTGAAGCAACTACTTTTACAAACAACGGCTGATGGATTAGTTGATATCAGAGG 1719  
QY 1240 TTGAAGAAGGTACATACACTATCTAGTTTGAGAAAAGGCTCCCTTAGGTTACAATTTGTTA 1299  
Db 1720 CTTAAATACGGTACCTATTATTTAGAGAAACTGTGTAGCTCCTGATGATTATGCTTTGTTA 1779  
QY 1300 GATTAACCTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359  
Db 1780 ACAATCGGATTTGAATTTTGGTCAATGAACATCATATGGCACAA-----CAGAAAAC 1833  
QY 1360 CTTTATAGTTAAACCAACTGTTGAAATACAAAGGTACTGAGTTGCCCTTCAACAGGTGGT 1419  
Db 1834 CTAGTTTCACAGAAAAGTACCAACAACAAGGTACCTTACCTTCAACAGGTGGC 1893  
QY 1420 ATTGGTACAAATTTTCTACATATATAGTGCAATTTTAGTAAATAGGAGCAGGTATCGTG 1479  
Db 1894 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTATGAGGAGTCTAC 1953  
QY 1480 CTTGTTGCTCGTCTGCTGTTTACGTTCTTAA 1509  
Db 1954 TTTGCTAGACGTAGAAAAGAAATGCTTAA 1983  
RESULT 7  
ABK11591  
ID ABK11591 standard; DNA; 2199 BP.  
XX AC ABK11591;  
XX DT 05-JUN-2002 (first entry)  
XX DE E. faecalis genomic region containing an Emac/D homologue.  
XX EX Extracellular matrix adhesion; Emac/D; ds; group B streptococcus; GBS;  
XX KW DNA vaccine; Spbi; Spb2; Rib; Lmb; C5a-ase; C protein alpha antigen;  
XX KW neonatal bacterial infection.  
XX OS Enterococcus faecalis.  
XX PN WO200212294-A2.  
XX PD 14-FEB-2002.  
XX PF 08-AUG-2001; 2001WO-US024795.  
XX PR 08-AUG-2000; 2000US-00634341.  
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX (UTAH ) UNIV UTAH RES FOUND.  
XX Adderson E, Bohnsack J;  
XX WPI; 2002-257465/30.  
XX New streptococcal matrix adhesion (Emac) polypeptides, useful as vaccines,  
XX particularly for treating or preventing infections by virulent forms of  
XX streptococci.  
XX Example 3; Page 167-169; 177pp; English.  
XX The invention relates to isolated streptococcal polypeptides, which  
XX comprise the Streptococcus agalactiae extracellular matrix adhesion (Emac)  
XX polypeptides EmaA, EmaB, Emac, EmaD or EmaE, and their encoding nucleic  
XX acids. Also included are the a vaccine comprising an Ema protein, an anti  
XX -Ema antibody, a pharmaceutical composition comprising one or more Ema  
XX proteins plus an anti Ema antibody, optionally in combination with at  
XX least one antibody to a protein selected from Spbi and Spb2, Rib, Lmb,  
XX Csa-ase or C protein alpha antigen, an immortal cell line producing an  
XX anti-Ema monoclonal antibody, a vector comprising the Ema nucleic acid, a  
XX host cell transformed with the vector, a nucleic acid vaccine comprising  
XX the vector and Ema homologous proteins and their encoding nucleic acids

CC from additional bacterial species (*S. pneumoniae*, *S. pyogenes*,  
CC *Enterococcus faecalis* and *Corynebacterium diphtheriae*). The streptococcal  
CC polyptides are useful as vaccines, particularly for treating or  
CC preventing infections by virulent forms of streptococci, especially group  
CC B streptococci (GBS) the most common cause of serious bacterial disease  
CC in neonates. The present sequence is an *E. faecalis* genomic region  
CC containing an Emac/D homologue

Seq	Sequence	2199 BP;	813 A;	380 C;	401 G;	605 T;	0 U;	0 Other;
	Query Match	4.3%;	Score 64.8;	DB 6;	Length 2199;			
	Best Local Similarity	52.1%;	Pred. No. 0.00014;					
	Matches	172;	Conservative 0;	Mismatches 152;	Indels 6;	Gaps 1;		
QY	1180	ACGAACTCTAATGCAACAGAAATATACAAACGAGCAGATGTTATAATTACATTACAGGC	1239					
Db	275	ACAAAAGCTGAAGCAACTACTTTTACAAACAACGGCTGATGATTAGTTGATATCACAGG	334					
QY	1240	TTGAAAGAAAGGTACATACTATCTAGTTTGAGAAAAAGGCTCCCTTAGGTTTCAAAATTCGT	1299					
Db	335	CTTAAATACGGTACCTATTATTTAGAAAGAAACTGTAGCTCTGATGATTATGCTTGT	394					
QY	1300	GATAACTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAAATTCAGATAAC	1359					
Db	395	ACAAATCGGATTGAAATTTGTGGTCAATGAAACAATCATATGSCACAA-----CAGAAAC	448					
QY	1360	CTTTTAGTTAACCAACTGTTTGAAAAATACAAAGGTACTGAGTTGCCTTTCAACAGGTGGT	1419					
Db	449	CTAGTTTCACCCAGAAAAAGTACCAACAACAAACACAAGGTACCTTACCCTTCAACAGGTGGC	508					
QY	1420	ATTGGTACAACAATTTTCTCATTTATAGGTGSCAAATTTTAGTAAATAGGAGCAGGTATCGT	1479					
Db	509	AAAGGAATCTACGTTTACTTAGGAAGTGGCGCACTCTTGCTACTTATTGACGAGAGTCTAC	568					
QY	1480	CTTGTTGCTCGTCTCGTTCGTTTACGTTCTTAA	1509					
Db	569	TTTGCTAGACGTAGAAAAAGAAAAATGCTTAA	598					

RESULT 8	
AAAX12982	
ID	AAAX12982 standard; DNA; 15614 BP.
XX	
AC	AAAX12982;
XX	
DT	19-MAR-1999 (first entry)
XX	
DE	Enterococcus faecalis genome contig SEQ ID NO:45.
XX	
KW	Enterococcus faecalis; contig; detection; Enterococcal infection;
XX	vaccine; attenuation; computer readable medium; ds.
XX	
OS	Enterococcus faecalis.
XX	
PN	WO9850555-A2.
XX	
PD	12-NOV-1998.
XX	
PF	04-MAY-1998; 98WO-US008985.
XX	
PR	06-MAY-1997; 97US-0044031P.
PR	16-MAY-1997; 97US-0046655P.
PR	14-NOV-1997; 97US-0066009P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Kunsch CA, Dillon PJ, Barash SC;
XX	
DR	WPI; 1999-045171/04.

Claim 1; Page 419-427; 2084pp; English.

A computer readable medium has been developed which has recorded on it 982 nucleotide sequences isolated from the *Enterococcus faecalis* genome. AAX12938 to AAX13919 represent these nucleotide sequences which are primary nucleotide sequences, also known as contigs. The computer-based system can identify fragments of the *Enterococcus faecalis* genome with commercial importance. The products can be used to detect the presence of *Enterococcus faecalis* in samples. They can also be used for diagnosing *Enterococcal* infection in an animal and monitoring progression of disease, and for identifying agents which can be used to modulate the growth or pathogenicity of *Enterococcus faecalis*, or another related organism, *in vivo* or *in vitro*. In particular the polypeptides encoded by the *Enterococcus faecalis* nucleotide sequences can be used in vaccines to prevent or attenuate an *Enterococcal* infection

Sequence 15614 BP; 5535 A; 2567 C; 3169 G; 4326 T; 0 U; 17 Other;

SQ	Sequence	15614 BP;	5535 A;	2567 C;	3169 G;	4326 T;	0 U;	17 Other;	
Query Match	4.3%;	Score	64.8;	DB	2;	Length	15614;		
Best Local Similarity	52.1%;	Pred. No.	0.00021;						
Matches	172;	Conservative	0;	Mismatches	152;	Indels	6;	Gaps	1;
Qy	1180	ACGAAGCTAATGCAACAGCAATATACACAGAGCAGATGGTATTAATTTACCAATTACAGGC	1239						
Qy	1240	TTGAAAGAAGGTACACTACTATCTAGTTGAGAAAAAGGCTCCCTTAGGTTTCAATTTGTGTTA	1299						
Db	13784	CYTAAATACGGTACCTATTATTTGAAGAAACTGTAGTCTCTGATGATTGTCTTGTTA	13843						
Qy	1300	GATAACTCTCAGAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTTACAGATAAC	1359						
Qy	1360	CYTTTAGTTTAACCCAACTGTTGAAAATAACAAAGGTACTGAGTTGCCCTTCAACAGGTGGT	1419						
Db	13898	CTAGTTTTCACGAAAAGGTACCAACAACAACAAAGGTACCTTACCTTCAACAGGTGGC	13957						
Qy	1420	ATTGGTACAAACAATTTCTFACATTTATAGTGGCAATTTTATAGTAATAGGACAGGTATCTGTG	1479						
Db	13958	AAAGGAATCTAGTTTACTTTAGGAAGTGGCGAGCTTGTCTACTATTATTCAGGAGTCTAC	14017						
Qy	1480	CTTGTTGCTCGTGGTTCAGTTCTTAA	1509						
Db	14018	TTTGCTAGACGTAGAAAAGAAAATGCTTAA	14047						

RESULT 9	
ABS98777	
ID	ABS98777 standard; DNA; 15614 BP.
XX	
XX	
AC	ABS98777;
XX	
DT	18-DEC-2002 (first entry)
XX	
DE	Enterococcus faecalis contig sequence #45.
XX	
KW	Computer readable medium; Enterococcus faecalis; microbe; growth; pathogenicity; vaccine; resistance; Enterococcal infection; commercial; therapeutic; industrial; fermenting; sugar source; metabolite; vaccine; biochip technology; antibacterial; modulator of nucleic acid expression; contig; ds.
XX	
XX	
OS	Enterococcus faecalis.
XX	
XX	US2002120116-A1.
XX	
PD	29-AUG-2002.
XX	
PF	04-MAY-1998; 98US-00070927.
XX	
PR	04-MAY-1998; 98US-00070927.







AC ABQ39490;  
XX 12-JUL-2002 (first entry)  
DT  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 26081.  
DE  
XX Human, cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200218632-A2.  
XX  
XX 07-MAR-2002.  
XX  
XX 01-SEP-2001; 2001WO-EP010074.  
XX  
XX 01-SEP-2000; 2000DE-01043826.  
PR  
XX 05-SEP-2000; 2000DE-01044543.  
PR  
XX (EPIG-) EPIGENOMICS AG.  
PA  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
PI  
XX WPI; 2002-371829/40.  
DR  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful for  
XX diagnosis and prognosis, comprises selective hybridization of amplicons  
XX from chemically treated DNA.  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
XX genomic sample of DNA. The sample is treated chemically to convert  
XX cytosine (C) but not methylated C, to uracil, then part of the genomic  
XX DNA that contains the target C is amplified to form a labeled amplicon.  
XX The amplicon is hybridised to two classes, each with at least one member,  
XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
XX degree of hybridisation to both classes is determined from the label on  
XX the amplicon. From the ratio of labels hybridised to the two classes of  
XX oligomers, the degree of methylation is calculated. The method is used:  
XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
XX and of a wide range of diseases, e.g. cancer, disorders of the central  
XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
XX particularly by detecting mutations or single nucleotide polymorphisms  
XX (SNP's); and (ii) for differentiation of cell or tissue types and for  
XX investigating cell differentiation. The method allows the methylation  
XX status of many C residues to be determined simultaneously. ABQ13410-  
XX ABQ54121 represent genomic DNA sequences used to illustrate the method  
XX for determining the degree of cytosine methylation described in the  
XX disclosure of the invention  
XX  
XX Sequence 969 BP; 177 A; 54 C; 93 G; 645 T; 0 U; 0 Other;  
SQ  
Query Match 3.8%; Score 56.6; DB 6; Length 969;  
Best Local Similarity 42.5%; Pred. No. 0.0075;  
Matches 360; Conservative 0; Mismatches 484; Indels 3; Gaps 1;  
90 TGAGACAGGCAATATACAGTTCAAGTACTCAAAAAGGGCGCAACCTATTAAGCATATAA 149  
849 TAAACGAATAAATAACGAATAAATACTAATAAAGCAATAAATAAATAAATAA 790  
150 AGTTTTTTCATGCAGTAAGTATGCAATGATCTGATTCGAATAAAGTGCAGTTC 209  
789 TACTAATAAAGCAATAAATAAATACGAATAAATACTAATAAAGCAATAAATAACGA 730  
210 TTATTTTAATTCCTCAAGGTAAGAGAGCTGAGTATAAAGCTTCAACTGATTTTCTCT 269  
729 TAAATAACTATAAAGCAATAAATAAATACGAATAAATAAATACTAATAAAGCAATAA 670

QY 270 TTTTACGACAACTACTAATGGAGGGAGACATATGTAACTTAAAAAGATACCTCGGTGACG 329  
DB 669 AACGAATAAATACTAATAAAGCAATAAATAAAGCAATAAATAAATACTAATAAAGCA 610  
QY 330 AAATGAGATTGGACATGGGCTAAATCTATATCAGCTAATACTACACCACTTTCCACTGT 389  
DB 609 TAAAAATACGAATAAATAAATACTAATAAAGCAATAAATAAATACTAATAAAGCAATA 550  
QY 390 TACTGAGTCAATAATGATGGTACTGAGGTATTAATGTTTCCCAATATGATATTTATTA 449  
DB 549 AACGAATAAATAAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAAT 490  
QY 450 TGTTCCTAGCCTGTTAATAATGGAGCTGTAATATGTTTATCTGTACTCAATGTC 509  
DB 489 TAACGAAACGAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAAATAA 430  
QY 510 TACTATTTCATGAAGAATACTGATGCGACATGGGAGATGGTGGTGGAAAACTGTAGA 569  
DB 429 AAATACTAATAAAGCAATAAATACTAATAAAGCAATAAATAAATACTAATAAAGCA 370  
QY 570 TCAAAAAACGTTACTCGGTGGTGATACAGTCAATATATCTTATTATTAAGAAATGCAGT 629  
DB 369 AAATAAAAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAAATAA 310  
QY 630 CAATTATCATGGTACAGAAAAAGTGTATCAATATGTTTATAAGGATATCTATGCCATCTGC 689  
DB 309 GAATACGAATACGAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAA 250  
QY 690 TTCTGTAGTGTGTTGACGAAGGGTCTTATGAAGTAACTATTACTGATGGATCAGGAA 749  
DB 249 TACTAATAAAGCAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAA 190  
QY 750 T---ATTACAACTCTAAGTTCGGAAGGCACTGGGAAGTATTAACCTGTTTGA 806  
DB 189 TAAAAATACTAATAAAGCAATAAATAAATAAATAAATAAATAAATAAATAAATAA 130  
QY 807 GGAATAAATAAATTTTCAGGATTAATTCCTGGTGGGCGACTACCAATATCTCCAAACCGGAA 866  
DB 129 TAAAAATAAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAAATAA 70  
QY 867 TACTCAAAATGGAGCTAATGATGATCTTTTATTAAGGGAATAAATAAATAAATAA 926  
DB 69 AACGACGCAATAAATAAATAAATACTAATAAAGCAATAAATAAATAAATAAATAA 10  
QY 927 TTATACA 933  
DB 9 TAATAAA 3  
RESULT 13  
ABQ39491  
ID ABQ39491 standard; DNA; 969 BP.  
XX  
XX ABQ39491;  
XX  
XX 12-JUL-2002 (first entry)  
DT  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 26082.  
DE  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
KW SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200218632-A2.  
XX  
XX 07-MAR-2002.  
XX  
XX 01-SEP-2001; 2001WO-EP010074.  
PF  
XX 01-SEP-2000; 2000DE-01043826.  
PR

PR 05-SEP-2000; 2000DE-01044543.  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.  
DR  
XX  
XX  
PT Determining the degree of cytosine methylation in genomic DNA, useful for  
PT diagnosis and prognosis, comprises selective hybridization of amplicons  
PT from chemically treated DNA.  
XX  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.  
XX  
XX This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ3410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX  
SQ Sequence 969 BP; 645 A; 93 C; 54 G; 177 T; 0 U; 0 Other;

Query Match 3.8%; Score 56.6; DB 6; Length 969;  
Best Local Similarity 42.5%; Pred. No. 0.0075;  
Matches 360; Conservative 0; Mismatches 484; Indels 3; Gaps 1;

QY 90 TCAGCAGGCAATACGATTCAGATCTCAAGATACCAAAAGGCGCACTATAAGCATATAA 149  
DB 121 TAAACGAAATAAAATACGAATAAAATACGAATAAAATACGAATAAAATACGAATAAA 180  
QY 150 AGTTTTTGATGAGAAATAGATAATGCAAAATGTATCTGATTCGAATAAAGATGGAGCTTC 209  
DB 181 TACTAATAAACGATAAAATACGATAAAATACGATAAAATACGATAAAATACGATAAA 240  
QY 210 TTATTTAATTCCTCAAGGTAAGAAGCTGAGTATTAAGCTTCAACTGATTTTAAATCTCT 269  
DB 241 TAAAAATACTAATAAAACGATAAAATACGAATAAAATACGAATAAAATACGAATAAA 300  
QY 270 TTTTACGCAACTACTTAATGGAGGAGACATATGTAACATAAAAGATCTGCGTCTAGC 329  
DB 301 AACGAATAAAATACTAATAAAACGATAAAATACGAATAAAATACGAATAAAATACGA 360  
QY 330 AAATGAGATTCGACATGGCTGATTAATCTATATCAGTATACATACACGATTTCCACTGT 389  
DB 361 TAAAAATACGATAAAATACGATAAAATACGATAAAATACGATAAAATACGATAAA 420  
QY 390 TACTGAGTCAATAATGATGGTACTGAGGTTTATTAATGTTTCCCAATATGGATATTATTA 449  
DB 421 AACGAATAAAATACGATAAAATACGATAAAATACGATAAAATACGATAAAATACGATA 480  
QY 450 TGTTCCTGACCTGTTAATAATGGAGCTGTAATTATGTTTACATCTGTAACCTCAATGCC 509  
DB 481 TAAACGAAACGATAAAATACGATAAAATACGATAAAATACGATAAAATACGATAAA 540  
QY 510 TACTATTTCATGAAAGAAATACGATGCGACATGGGAGATGTTGTTGCGAAACCTGTAGA 569  
DB 541 AAATACTAATAAAACGATAACGATAACGATAACGATAACGATAACGATAACGATAACG 600

QY 570 TCAAAAACGATCTCGGTGGTGATACAGTCAAAATATATCTATTACTTATTAAGAATGCAGT 629  
DB 601 AAATAAAAATAAAAATATCTAATAAAACGAATAAAAATACGAATAAAAATATCTAATAAAC 660  
QY 630 CAATTATCATGGTACAGAAAAAGTGATCAATATGTTTATTAAGGATATCTATGCCATCTGC 689  
DB 661 GAATACGANTACGAATACAAATAAAAATACGATAAAACGAAACGAAATAAAAATAAAA 720  
QY 690 TTCTGTAGTTGATTTGAACGAAGGCTCTTATGAAGTAACTTATCTGATGGATCAGGAA 749  
DB 721 TACTAATAAAACGAATAAAAATACGAATAAAAATACGATAAAACGAATAAAAATACGAA 780  
QY 750 T---ATTACAACCTCTAAGCTCGGAAACGAACCTCGGAAAGTATTAACCTGTTTGA 806  
DB 781 TAAAAATACTAATAAAACGAATAAAAATACGAATAAAAATACGAATAAAAATACGAACGAA 840  
QY 807 GGAATAAATAAATTTTCCAGATTTACTATTTCCTGGGAGCTACCAATCTCCCAACCGGAA 866  
DB 841 TAAAAATAAAATAAAATACGATAAAACGATAAAATAAAATAAAATAAAATAAAATAA 900  
QY 867 TACTCAAAATGGAGCTAATGATGACTTTTTTTTATAAGGGAATAAAATACAATCACAGTCAC 926  
DB 901 AACGAACGAAATAAAAATAAAATACGATAAAACGAATAAAAATACGAATAAAAATAC 960  
QY 927 TTATACA 933  
DB 961 TAATAAA 967

RESULT 14  
ABL28164  
ID ABL28164 standard; DNA; 2408 BP.  
XX  
AC ABL28164;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 35965.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 35965; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the

CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 2408 BP; 724 A; 504 C; 437 G; 743 T; 0 U; 0 Other;
	Query Match            3.7%;   Score 55.4;   DB 4;   Length 2408;
	Best Local Similarity   49.9%;   Pred. No. 0.017;
	Matches 195; Conservative   0; Mismatches 191; Indels   5; Gaps   2;
QY	20   AATCGCTGTTAGTGGCGAGTTTAGCATTTTCGTATGGCTGTGTATCAC--CAGTTACGCCGAT 77
DB	963   ATTCAGTTTTAGCTGTGTAGCTTTAGCTATAGCTTTGGCGTATAACCACACAGATACAGATAC 1022
QY	78   AGCTTTTGCCGCTGAGACAGGACAATTACAGTTTCAAGTACTCAAAGAAGCGCAACCTA 137
DB	1023   AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1082
QY	138   TAAAGCATATAAAGTTTTTCATGCAGAAAAATAGATAAATGCAAAATGTTATCTGATTCGAATAA 197
DB	1083   AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATAC 1142
QY	198   AGATGGAGCTTCTTTATTAAATTCCTCAAGSTAAGAAGCTGAGTATTAAGCTTCAACTGA 257
DB	1143   AGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATA 1199
QY	258   TTTTAAATTCCTTTTACGCAACTACTAATGGGGGAGAACATATCTGAATAAAAAAGA 317
DB	1200   TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGA 1259
QY	318   TACTGCGTCAGCAAAATCAGAAWTGGCGCATGGCGTAAATCTATATCGCTAATCTACTACACC 377
DB	1260   TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGA 1319
QY	378   AGTTTCOACTGTTTACTGAGTCAAAATATGAT 408
DB	1320   TACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGATACAGAT 1350

RESULT 15	
ABLS6203/c	
ID	ABL56203 standard; DNA; 32392 BP.
XX	
XX	
AC	ABL56203;
XX	
XX	01-JUL-2002 (first entry)
DT	
XX	
DE	AmEPV genome fragment#5.
XX	
XX	AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW	genetic deficiency disorder; ds.
KW	
XX	
OS	Amsacta moorei entomopoxvirus.
XX	
XX	WO200212526-A2.
FN	
XX	
PD	14-FEB-2002.
XX	
XX	10-AUG-2001; 2001WO-US025287.
PF	
XX	
XX	10-AUG-2000; 2000US-0224479P.
PR	
PR	14-SEP-2000; 2000US-00662254.
XX	
XX	(UYFL ) UNIV FLORIDA.
PA	
XX	
XX	Moyer RW, Li Y, Bawden AL;
PI	
XX	
XX	WPI; 2002-227161/28.
DR	
XX	
XX	Novel recombinant entomopox virus vector useful for delivering
PT	polynucleotide encoding protein to vertebrate cell, comprises
PT	polynucleotide encoding protein operably linked with heterologous
PT	promoter sequence.
XX	
XX	

PS Disclosure; Page 226-242; 326pp; English.

XX The invention relates to a recombinant entomopox virus (EPV) vector, comprising a polynucleotide encoding a protein operably linked with a heterologous promoter sequence. The invention also concerns methods for providing gene therapy for genetic deficiency disorders. Vectors of the invention are useful for delivering a polynucleotide encoding a protein to a vertebrate cell preferably a mammalian cell, such as a human cell. The vector is introduced into the vertebrate cell by infection in vitro or in vivo. The vector is useful for the delivery and expression of biologically useful proteins in gene therapy protocols, and for delivering large DNA segments for engineering of vertebrate cells. Polynucleotides of the invention have applications in techniques such as their use as insertion sites for foreign genes of interest, hybridisation probes, for chromosome and gene mapping, in PCR technologies, and in the production of sense or antisense nucleic acids. Vectors of the invention provide for stable integration and expression of heterologous DNA in host cells, and are adapted for accepting large heterologous polynucleotide inserts which can be delivered in an infected or transformed cell and expressed in a stable fraction. The current sequence represents a fragment of the genome of the genus B entomopoxvirus from amsacta moorei (AmEPV)

XX Sequence 32392 BP; 13748 A; 2577 C; 2550 G; 13517 T; 0 U; 0 Other;

SQ

Query Match 3.7%; Score 55.2; DB 6; Length 32392;  
Best Local Similarity 44.0%; Pred. No. 0.034;  
Matches 329; Conservative 0; Mismatches 413; Indels 6; Gaps 2;

479 TAATTATGGTTACATCTGTAACCTCAAATGCTACTATTCATGTGAAGAAGATACTGATCGCA 538  
|||||  
1602 TAATAATATAAATGGATATATTACTAATTGCCTTAAACTATCATATTTTTTCAGATCATA 1543  
|||||  
539 CATGGGAGATGCTGCTGGAAAACTGTAGATCAAAAACGTACTCGGTGGTGATACAG 598  
|||||  
1542 ACATATATGTTGAATGAAGAAAAATTAGAGNTTTTAATAATAATATCTCGTGTAA 1483  
|||||  
599 TCAAAATATACATTACTTATAAGAANTGCCAGTCAATTTATCATGGTACAGAAAAAGTGTATC 658  
|||||  
1482 TTATAATTTATATTA--ACAATAATACTAAAAAAAATAATACTACAATAATAGATAAAT 1426  
|||||  
659 AATATGTTATAAGGATACCTNCCATCTGCTTCTGTAGTTGATTTGAACGAAGGGTCTT 718  
|||||  
1425 ATAATGAAATAGAAAAAATAAATAATTAATAAGAAATAGTGAATAGTAATAG--TG 1369  
|||||  
719 ATGAAGTAACTATTACTGTATGGATCAGGAATATTAACAATCTTAACCTCAAGTTCGGAAA 778  
|||||  
1368 ATTCCGACAGATCTAACTAATAAATAAGCAATATATATTTCAAATCAGTATATAGTAAAA 1309  
|||||  
779 AGCAACTGGGAAGTATPAACCTGTTAGAGAAAATAATAATTTTTCACGATTTACTATTCCGT 838  
|||||  
1308 CAGAAGAAGGAGTAATTCCTGTTATTAAACGTGATCTCTGATAATAATTAATAAAAAATAGTT 1249  
|||||  
839 GGCAGCTACCATPACTCCAACCGGAATACTCAAAATGGAGCTTAATGATGACATCTTTTTT 898  
|||||  
1248 TAGCAGAAGATATATCTTTAAAAGAGATGTTATAAAGAAGATTAATAATATAGATTTAT 1189  
|||||  
899 ATAAGGGAATAAATPACAATCACAGTCACTTATACAGGAGTATTAAAGAGTGGACCTAAAC 958  
|||||  
1188 ATGAGAAAAAATAATAATGTTAAATAACAATAGAAAATAATATATATCAAAAAAATAT 1129  
|||||  
959 CAGGTTCAGCTGATTTACCAGAAATAACAACATTTGCGACCATCAACCCCATAACTAGCA 1018  
|||||  
1128 TATCTTTTGATATAATAATAAAAAAATAATAATAATAGTGTGGTAAATTCATGAGAA 1069  
|||||  
1019 ATGATGACCCAGGTCAAAAAGTAAACGTGAGGGATGTCAAATTTACTATAAAAAAATTTG 1078  
|||||  
1068 TTATTAACGATGTATTAATTTGAAAGCAAAAATAACGAGAAAAAGACTATTCAAATACATTC 1009  
|||||  
1079 ATGGTTCCAAAAAGCTTCAATCAAGGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTC 1138  
|||||  
1008 CTATATATGAAAAAGTTCAAATAGAAATAATATATTTATTTTATTCGATTAATAAAGAGTA 949  
|||||

[illegible]





XX PR 05-NOV-1998; 98US-0107131P.  
XX PA (HOFF/) HOFFMAN S.  
XX PA (CARU/) CARUCCI D.  
XX PA (GARD/) GARDNER M.  
XX PA (VENT/) VENTER J C.  
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX DR WPI; 2000-365347/31.  
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
XX PT diagnosis of P.falciparum infection.  
XX PS Disclosure; Page 493-495; 577pp; English.  
XX CC The present invention describes proteins and their fragments (I) encoded  
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
XX CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
XX CC vaccines against P. falciparum infection comprising (I) or (II). (I) and  
XX CC (II) are useful for the development of vaccines against P. falciparum  
XX CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
XX CC immunogens comprising the sequences of (I), are useful in the detection  
XX CC of infection with P. falciparum. Furthermore, (I) (especially when they  
XX CC are rifins or secreted or membrane proteins) can aid the identification  
XX CC of drugs to treat or prevent P. falciparum infection, or they can be used  
XX CC to identify drug resistance in P. falciparum. Sequencing of the  
XX CC Plasmodium chromosome 2 and the subsequent identification of proteins  
XX CC encoded by it will help to expand our understanding of parasite biology,  
XX CC a process hampered by the complexity of the parasitic lifecycle, and  
XX CC provide new targets for vaccine and drug development. Parasite resistance  
XX CC to drugs and mosquito resistance to insecticides have led to a resurgence  
XX CC of malaria in many parts of the world, and there is a pressing need for  
XX CC vaccines and new drugs. AAAY0078 to AAA70287 and AAB18144 to AAB18352  
XX CC represent nucleotide and protein sequences given in the present  
XX CC invention, but which are not specifically mentioned within the  
XX CC specification  
XX SQ Sequence 6033 BP; 3019 A; 437 C; 707 G; 1870 T; 0 U; 0 Other;  
Query Match 3.5%; Score 53; DB 3; Length 6033;  
Best Local Similarity 43.6%; Pred. No. 0.072;  
Matches 236; Conservative 0; Mismatches 305; Indels 0; Gaps 0;  
390 TACTGAGTCAAAATGATGCTAGCTAGGTTATTAATGTTTCCCAATATGATATTTA 449  
978 TATAGATGTAATAATAAATAATGATGATGATAATAATGATGATAATAATAATA 1037  
450 TGTTCCTAGCAGCTTAATAATGAGCTGTAATTATGTTTACATCTGTAACCTCAAATGC 509  
1038 TAATGATGATAATAATAATAATAATAATGATGATAATAATAATGATGATAATAATA 1097  
510 TACTATTTCATGAAAGAAATCTGATCGGACATCGGGAGATGGTGGGAAAACTGTAGA 569  
1098 TAATAATAATAATAATGATGATAATAATAATAATAATAATAATAATAATAATAATA 1157  
570 TCAAAAACCTACTCGTGGTGGTATGATGATGATGATGATGATGATGATGATGATG 629  
1158 TTATTATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1217  
630 CAATTATCATGGTACAGAAAGTGTATCAATATGTTTATAAGAGATCATGTCATCTGC 689  
1218 TATGTTAAATATACAAAATATTATTACAAAACATTTTAAAGAGATAAAAAAATATTTA 1277  
690 TTCTGTAGTGAATTTGAACGAAGGCTTTATGAAGTAACATTTACTGATGATCAGGAA 749  
1278 CACAAATAATAATAAAGAAAGAAATTTATTTTCCCACTAATGGAACATTTGTATATGATA 1337  
750 TATTACAACTCTCACTCAAGTTTCGGAAAAAGCACTGGGAAGTATACCTGTTAGAGGA 809  
1338 AAAAAAATCTTTGATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1397

QY 810 AAATAATAATTTCCAGATTACTATTCCGTGGCGAGCTACCAATATCTCCAAACCGGAAATAC 869  
Db 1398 CGATAAAATTAATTAAGTTTAATAATAATAATAATAATAATAATAATAATAATAATAATA 1457  
QY 870 TCAAAATGGAGCTAATGATGACTTTTTTATAAAGGAATAAATAACAATCAACAGTCACCTTA 929  
Db 1458 TTGCAAAATTAACGGCATATGTAATTTTGTGTATATATTAGGAATAAGTGAATATATTTA 1517  
QY 930 T 930  
Db 1518 T 1518  
RESULT 21  
ABN67916  
ID ABN67916 standard; DNA; 2703 BP.  
XX AC ABN67916;  
XX DT 01-JUL-2002 (first entry)  
XX DE Streptococcus polynucleotide SEQ ID NO 3745.  
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX OS Streptococcus agalactiae.  
XX PN WO200234771-A2.  
XX PD 02-MAY-2002.  
XX PF 29-OCT-2001; 2001WO-GB004789.  
XX PR 27-OCT-2000; 2000GB-00026333.  
XX PR 24-NOV-2000; 2000GB-00028727.  
XX PR 07-MAR-2001; 2001GB-00005640.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Telford J, Massignani V, Margarit Y Rosl, Grandi G, Fraser C;  
XX PI Tettelin H;  
XX DR WPI: 2002-352536/38.  
XX DR P-PSDB; ABP27285.  
XX PT New Streptococcus protein for the treatment or prevention of infection or  
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for  
XX PT detecting a compound that binds to the protein.  
XX PS Claim 7; Page 3529; 4525pp; English.  
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
XX CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
XX CC the specification. The proteins have antibacterial and antiinflammatory  
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX CC antibodies that bind (I) are used in the manufacture of medicaments for  
XX CC the treatment or prevention of infection or disease caused by  
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX CC biological sample. (I) is used to determine whether a compound binds to  
XX CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX CC used as a vaccine or diagnostic composition. The disease caused by  
XX CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX CC acid encoding (I) may be used to recombinantly produce (I) and may be  
XX CC used in gene therapy. Antibodies to (I) are used for affinity  
XX CC chromatography, immunoassays, and distinguishing/identifying  
XX CC Streptococcus proteins



SQ	Sequence	2703 BP; 1074 A; 376 C; 500 G; 753 T; 0 U; 0 Other;	
	Query Match	3.4%; Score 51.2; DB 6; Length 2703;	
	Best Local Similarity	50.4%; Pred. No. 0.15;	
	Matches 125; Conservative	0; Mismatches 123; Indels 0; Gaps 0;	
QY	1080	TGGTTCCCAAGGCTTCATTACAGGTCCTATATTTGTTTAAAGATGCTACGGGTCA	1139
DB	2214	TGAATTTATAAAGTTGATAAGATAATAAAGTTGCTTCTCAAAGAGGCTACGTTTGA	2273
QY	1140	ATTTCCTAAACTTTAACGATACAAATAAAGCTTGAATGGGGCAGAGAGCTTAATGCAACAGA	1199
DB	2274	ACTTCAAGAAATTAATGAAGATTATTAACCTTTATTTACCAATAAAAATAATAATTCAA	2333
QY	1200	ATATCAACAGAGCAGATGGTATTAATTACCATTTACAGCTTTGAAAGAGGTACATCTA	1259
DB	2334	AGTAGTGGCGGAGAAACGGCAAAATTTCTTACAAAGATTTGAAAGATGGCAAAATATCA	2393
QY	1260	TCTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATTAATCTCAGAAGGTTAT	1319
DB	2394	GTTAATAGAGCAGTTTCGCCGGAGGATTTATCAAAAAATTTACTAATAAACCAATTTTAAC	2453
QY	1320	TTTAGGAG 1327	
DB	2454	TTTGAAG 2461	
RESULT 22			
ABN71527_13/c			
Continuation (14 of 22) of ABN71527 from base 1300001 (Streptococcus polynucleotide SEQ			
WP Sequence split into 22 fragments LOCUS ABN71527 Accession Abn71527			
WP	Fragment Name	Begin	End
WP	ABN71527_00	1	110000
WP	ABN71527_01	100001	210000
WP	ABN71527_02	200001	310000
WP	ABN71527_03	300001	410000
WP	ABN71527_04	400001	510000
WP	ABN71527_05	500001	610000
WP	ABN71527_06	600001	710000
WP	ABN71527_07	700001	810000
WP	ABN71527_08	800001	910000
WP	ABN71527_09	900001	1010000
WP	ABN71527_10	1000001	1110000
WP	ABN71527_11	1100001	1210000
WP	ABN71527_12	1200001	1310000
WP	ABN71527_13	1300001	1410000
WP	ABN71527_14	1400001	1510000
WP	ABN71527_15	1500001	1610000
WP	ABN71527_16	1600001	1710000
WP	ABN71527_17	1700001	1810000
WP	ABN71527_18	1800001	1910000
WP	ABN71527_19	1900001	2010000
WP	ABN71527_20	2000001	2110000
WP	ABN71527_21	2100001	2155561
	Query Match	3.4%; Score 51.2; DB 6; Length 110000;	
	Best Local Similarity	50.4%; Pred. No. 0.35;	
	Matches 125; Conservative	0; Mismatches 123; Indels 0; Gaps 0;	
QY	1080	TGGTTCCCAAGGCTTCATTACAGGTCCTATATTTGTTTAAAGATGCTACGGGTCA	1139
DB	84684	TGAATTTATAAAGTTGATAAGATAATAAAGTTGCTTCTCAAAGAGGCTACGTTTGA	84625
QY	1140	ATTTCCTAAACTTTAACGATACAAATAAAGCTTGAATGGGGCAGAGAGCTTAATGCAACAGA	1199
DB	84624	ACTTCAAGAAATTAATGAAGATTATTAACCTTTATTTACCAATAAAAATAATAATTCAA	84565
QY	1200	ATATCAACAGAGCAGATGGTATTAATTACCATTTACAGCTTTGAAAGAGGTACATCTA	1259
DB	84564	AGTAGTGGCGGAGAAACGGCAAAATTTCTTACAAAGATTTGAAAGATGGCAAAATATCA	84505
QY	1260	TCTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATTAATCTCAGAAGGTTAT	1319

DB	84504	GTTAATAGACAGCTTTCGCCGGAGGATTATCAAAAAATTTACTAATAAACCAATTTTAAC	84445
QY	1320	TTTAGGAG 1327	
DB	84444	TTTGAAG 84437	
RESULT 23			
AAAX99556/c			
ID	AAAX99556	standard; DNA; 14066 BP.	
XX	AAAX99556;		
AC	AAAX99556;		
XX	05-OCT-1999	(first entry)	
DT	05-OCT-1999	(first entry)	
DE		Nucleic acid sequence from U. urealyticum.	
XX			
KW		Ureaplasma urealyticum; nucleic acid detection; infection; pathogen;	
KW		human urogenital tract; pregnancy; neonatal disease; drug therapy;	
XX		suppurative arthritis; ss.	
OS		Ureaplasma urealyticum.	
XX	WO9939007-A1.		
PN	WO9939007-A1.		
XX	05-AUG-1999.		
PD	05-AUG-1999.		
PF	29-JAN-1999;	99WO-US001972.	
PR	30-JAN-1998;	98US-0073189P.	
XX	(UABR-) UAB RES FOUND.		
PA	Cassell GH, Chen EY, Glass JS, Glass JI, Heiner CR, Lefkowitz E;		
XX	WPI; 1999-469343/39.		
DR	Detection of Ureaplasma urealyticum using novel genes, probes and		
XX	primers.		
PS	Claim 1; Page 41-46; 110pp; English.		
CC	The present invention provides methods for the detection and diagnosis of		
CC	Ureaplasma urealyticum infection. It provides novel genes (AAAX99501-681)		
CC	that can be used as a source of primers and probes for the detection and/		
CC	or quantification of U. urealyticum in a biological sample. The probes		
CC	that can be used in the method of the invention by forming target:probe		
CC	complex is complementary to a region selected from one of the 181		
CC	nucleic acid sequences (AAAX99501-681). U. urealyticum is an opportunistic		
CC	pathogen of the human urogenital tract that is a significant cause of		
CC	adverse pregnancy outcome, neonatal disease, and suppurative arthritis.		
CC	As the infections are commonly asymptomatic, it is important to have		
CC	specific and sensitive methods for detecting their presence in a patient.		
CC	Also, as the pathogen has no current antibiotic directed specifically		
CC	against it, it would be advantageous to isolate and detect gene sequences		
CC	which are unique to it, and utilise these as a basis for diagnosis of U.		
CC	urealyticum infection as well as to develop new and improved drug		
CC	therapies. The present invention provides such novel polynucleotide		
CC	sequences (AAAX99501-681)		
XX	Sequence 14066 BP; 4508 A; 1702 C; 1539 G; 6317 T; 0 U; 0 Other;		
SQ	Sequence 14066 BP; 4508 A; 1702 C; 1539 G; 6317 T; 0 U; 0 Other;		
	Query Match	3.3%; Score 50.4; DB 2; Length 14066;	
	Best Local Similarity	42.8%; Pred. No. 0.33;	
	Matches 313; Conservative	0; Mismatches 416; Indels 3; Gaps 1;	
QY	435	ATATGGATATTTATTTATGTTTCTTAGCACTGTTAATAATGGAGCTGTAATATGTTTACATC	494
DB	5284	AAATGTTAGTTTCTAATGATTTCTTAAATACTACAACACAAACAGTTAATATTTGATTTGA	5225
QY	495	TGTAACCTCAAACTGCTACTATTTCATGAAAGCAATCTGATCGACATGGGAGATGGTGG	554
DB	5224	TGGAAATTCACGAACCTTGAATTAATAAAAAAATTCATTTGGTATATATCTTCAATGACGG	5165



Db 9084 TAATTAACACCAATGATTAGTATTAAGTTCAATTAATGAGCCAAATGCAAGAGCGAAAA 9143  
 Qy 795 TAACCTGTTAGAGGAAAAATAATTTTACGATTAATTTCCGCGGAGCTACCAATAC 854  
 Db 9144 TAATTTACAATCAACACAAATTAGATTATTAATGATCTGATTAATGTTTAAAGTAA 9203  
 Qy 855 TCAACCGGAATACTCAAAATGGAGCTAATGATGACTTTTTTAAAGGGAAATAAATAC 914  
 Db 9204 TGAAGAAGAAGCAATCATTAATTTATGTTGAGAAAT--TTATCTGTAAAGCAAAAGTTAA 9260  
 Qy 915 AATCAGTCACTTATACAGGAGTATTAAGAGTGGAGCTAAACACGTTCAAGCTGATT 974  
 Db 9261 AATTGAGGAGTCAAAATATTTTAAACAACATTCATTAATCTAGCAATTAACCAAGA 9320  
 Qy 975 ACCAGAAATACAAACATTCGACCATCAACCCCAATCTAGCAATGATGACCCAGGTCA 1034  
 Db 9321 TACTATTATACAGTATTACTTTTATTATTAACCAATTAAGCTGTACGAATTTGG 9380  
 Qy 1035 AAAAGTAACAGTGGGAGTGGTCAAAATTAATTAATAAAAAATTTGATGTTCCACAAAAGC 1094  
 Db 9381 AATAGATAATAGCAATAAAATATATAACAATAACAACAATATAAAATCCATTAAAAATTTGA 9440  
 Qy 1095 TTCATTACAAGGTCTATATTTGTTTAAAGAACTGCTACGGTCAATTTCTAAACTTTAA 1154  
 Db 9441 TAATTAATTTTAGTGTCATTGTCCTGTTTCAAAAGTGATCTAATAAAGCTATTACTTTTGA 9500  
 Qy 1155 CGATACAAATAA 1166  
 Db 9501 ATATAATAATA 9512

## RESULT 25

AAA70250  
 ID AAA70250 standard; DNA; 7143 BP.  
 AC AAA70250;

07-NOV-2000 (first entry)

Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:383.

Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoicide; infection; insecticide; ds.

Plasmodium falciparum.

WO200025728-A2.

11-MAY-2000.

05-NOV-1999; 99WO-US026796.

05-NOV-1998; 98US-0107131P.

(HOFF/) HOFFMAN S.

(CARU/) CARUCCI D.

(GARD/) GARDNER M.

(VENT/) VENTER J C.

Hoffman S, Carucci D, Gardner M, Venter JC;

WPI; 2000-365347/31.

Proteins encoded by chromosome 2 of the human malarial parasite.

Plasmodium falciparum, useful as antimalarial vaccines and in the diagnosis of P.falciparum infection.

Disclosure; Page 560-562; 577pp; English.

The present invention describes proteins and their fragments (I) encoded

by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

Also described are: (I) nucleotide sequences (II) encoding (I); and (2)

CC vaccines against P. falciparum infection comprising (I) or (II). (I) and  
 CC (II) are useful for the development of vaccines against P. falciparum  
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
 CC immunogens comprising the sequences of (I), are useful in the detection  
 CC of infection with P. falciparum. Furthermore, (I) (especially when they  
 CC are rifins or secreted or membrane proteins) can aid the identification  
 CC of drugs to treat or prevent P. falciparum infection, or they can be used  
 CC to identify drug resistance in P. falciparum. Sequencing of the  
 CC Plasmodium chromosome 2 and the subsequent identification of proteins  
 CC encoded by it will help to expand our understanding of parasite biology,  
 CC a process hampered by the complexity of the parasitic lifecycle, and  
 CC provide new targets for vaccine and drug development. Parasite resistance  
 CC to drugs and mosquito resistance to insecticides have led to a resurgence  
 CC of malaria in many parts of the world, and there is a pressing need for  
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352  
 CC represent nucleotide and protein sequences given in the present  
 CC invention, but which are not specifically mentioned within the  
 CC specification

XX SQ Sequence 7143 BP; 3304 A; 520 C; 1000 G; 2319 T; 0 U; 0 Other;

Query Match 3.3%; Score 50.2; DB 3; Length 7143;

Best Local Similarity 43.2%; Pred. No. 0.31;

Matches 293; Conservative 0; Mismatches 383; Indels 3; Gaps 1;

Qy 550 GGTGGTGAAGAACTGTAGATCAAAACAGTCTCGGTGCTGATACAGTCAATATAT 609

Db 4 GGTAAATACGAACAGAAAAGATATAAGTCACAAAGAGTATGATAAAAGTTTATAAATATC 63

Qy 610 ATTACTTATAAGATGCGATCAATATATCATGTCAGAGAAAAGTGTATCAATATGTTATA 669

Db 64 GAGTCTGCTGAAGAACATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 123

Qy 670 AAGGATATCTATGCCATCTGCTTCTGTAGTTCGATTTGAACGAGGCTCTTATGAAGTAACT 729

Db 124 ATCGATAATAGTAATTAATTTGTAATTAATAGTAATTAATAGTAATTAATAGTAATTAAT 183

Qy 730 ATTACTGATGGATCAGGGAATATTACAACTCTAACTCAAGGTTCCGAAAAAGCAACTCGG 789

Db 184 AATAATAATAATAACAATATTGTTAGAAATAATAATAATAATAATAATAATAATAATAATA 243

Qy 790 AAGTATAACCTGTTAGAGGAAAATAATAATTTTCAGGATTAATTCCTGGGCGAGTACC 849

Db 244 AAGAATGTCTATTAATTAATGAGGATGATGATATAAATAAATAAATAAATAAATAAATAA 303

Qy 850 AATACTCCACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTATAAAGGGAATA 909

Db 304 TTGTGTAATATATTTTATGAAAATTTATTTTAAAACCTTTTAAATTAATGATGTG 363

Qy 910 AATAAATCAAGTCACCTTATACAGGAGTATTAAGAGTGGAGCTAAACCCAGTTCACT 969

Db 364 AGTAATAATAAGTTATAAATATTATTGAAACAAAAGGAGGAGATGAAAGGAATGCAGAT 423

Qy 970 GATTTACCAGAAAATACAAACATTTGCGACCATCAACCCCAATCTAGCAATGATGACCCA 1029

Db 424 AATAATTTTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 483

Qy 1030 GGTCAAAAAGTAAACAGTGGAGGATGGTCAAAATTAATTAATAAATAAATAAATAAATAA 1086

Db 484 AATGTAATGAAATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 543

Qy 1087 ACAAAGCTTCATTTACAAGGTGCTATATTTGTTTTTAAAGAAATGCTACGGGTCAATTTCTA 1146

Db 544 ACAAATGTAAATAGAAAATGCTAAATTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 603

Qy 1147 AACTTTAAGTACAAATAAGTTGAATGGGACAGAGTAATGCACAGAGATATACATA 1206

Db 604 AAGATATATTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 663

Qy 1207 ACAGGAGCAGATGGTATAA 1225

Db 664 AAGATATATTGTAATATAA 682

[illegible]

```
Db      1526 CTTAATACGGTACCTATTATTATTAGAGAACTGTAGCTCCTCGATGATTATGCTTGTTA 1585
Qy      1300 GATAA 1304
Db      1586 ACAAA 1590

RESULT 28
ABN98094
ID      ABN98094 standard; DNA; 1687 BP.
AC      ABN98094;
XX
DT      05-AUG-2002 (first entry)
XX
DE      E faecalis EF058 gene fragment.
XX
KW      Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
gene; ds.
XX
OS      Enterococcus faecalis.
XX
PN      US2002045737-A1.
XX
PD      18-APR-2002.
XX
PF      04-MAY-1998; 98US-00071035.
XX
PR      04-MAY-1998; 98US-00071035.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR      WPI: 2002-425450/45.
DR      P-PSDB; ABP43338.
XX
XX      New genes and polypeptides from Enterococcus faecalis, useful as vaccines
PT      for preventing, treating or attenuating an infection caused by a member
PT      of the Enterococcus genus in an animal, particularly E. faecalis.
XX
PS      Claim 1; Page 93-94; 255pp; English.
XX
XX      The present invention provides the protein and coding sequences of a
CC      number of polypeptides from Enterococcus faecalis. The proteins can be
CC      used as vaccines for preventing or attenuating an infection caused by a
CC      member of the Enterococcus genus in an animal, particularly E. faecalis.
CC      The polynucleotide is also useful for preventing or treating E. faecalis
CC      infection. The present sequence is a coding sequence of the invention
XX
SQ      Sequence 1687 BP; 632 A; 284 C; 360 G; 411 T; 0 U; 0 Other;

Query Match
Best Local Similarity 3.3%; Score 49.8; DB 6; Length 1687;
Matches 78; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy      1180 ACAGAAGCTTAATGCAACAGATATACACAGGAGCAGATGGTATTAATACCATACAGGC 1239
Db      1466 ACAAAAGCTGAGCAACTACTTTTACACACGGCTGATGATGATGATACACAGG 1525
Qy      1240 TTGAAGAAGGTACATACATCTAGCTTGAGAAAAAGGTCCTTAGGTTACAATTGTTA 1299
Db      1526 CTTAATACGGTACCTATTATTATTAGAGAACTGTAGCTCCTCGATGATTATGCTTGTTA 1585
Qy      1300 GATAA 1304
Db      1586 ACAAA 1590

RESULT 29
ACA88058
ID      ACA88058 standard; DNA; 1687 BP.
```

```
XX      ACA88058;
XX
DT      07-JUL-2003 (first entry)
XX
DE      E. faecalis novel gene #110.
XX
KW      Gene; ds; endocarditis; bacteraemia; urinary tract infection; UTI;
intraabdominal infection; soft tissue infection; neonatal sepsis;
vaccine.
XX
OS      Enterococcus faecalis.
XX
PN      US2003017495-A1.
XX
PD      23-JAN-2003.
XX
PF      29-JUL-2002; 2002US-00206576.
XX
PR      06-MAY-1997; 97US-0044031P.
PR      14-MAY-1997; 97US-0046655P.
PR      16-NOV-1997; 97US-0066009P.
PR      04-MAY-1998; 98US-00071035.
XX
PA      (HUMA-) HUMAN GENOME SCI INC.
XX
PI      Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR      WPI: 2003-416890/39.
DR      P-PSDB; ABU88366.
XX
XX      New nucleic acid molecules and polypeptides from Enterococcus faecalis,
PT      useful as vaccines for preventing or attenuating an enterococcal
PT      infection in an animal, or for identifying Enterococcus faecalis in
PT      biological samples.
XX
PS      Claim 1; Page; 40pp; English.
XX
XX      The invention relates to a new isolated nucleic acid molecule comprising
CC      a polynucleotide isolated from Enterococcus faecalis appearing as
CC      ACA87949-ACA88196 (or sequences complementary to them or 95% identical to
CC      them). Also included are the proteins encoded by the above nucleic acids,
CC      making a recombinant vector (comprising inserting the isolated nucleic
CC      acid molecule cited above into a vector), a host cell comprising the
CC      vector (used to produce the protein), an isolated antibody specific for
CC      the polypeptides, a hybridoma that produces the antibody, an isolated
CC      polypeptide antigen comprising an amino acid sequence of an Enterococcus
CC      faecalis epitope listed in the specification, a vaccine comprising one or
CC      more E. faecalis polypeptides (and a pharmaceutical diluent, carrier or
CC      excipient) where the polypeptide elicits protective antibodies in an
CC      animal to a member of the genus Enterococcus; preventing or attenuating
CC      an infection caused by a member of the genus Enterococcus in an animal
CC      comprising administering to the animal the polypeptide and detecting
CC      Enterococcus nucleic acids in a biological sample. The E. faecalis
CC      nucleic acid molecules and polypeptides are useful as vaccines for
CC      preventing or attenuating an enterococcal infection in an animal (e.g.
CC      endocarditis, bacteraemia, urinary tract infection (UTI), intraabdominal
CC      infection, soft tissue infection and neonatal sepsis). The polypeptides
CC      are also useful for detecting Enterococcus aureus in immunoassays, as
CC      epitope tags, as molecular weight markers, or for generating antibodies
CC      that specifically bind E. faecalis polypeptides. The nucleic acid
CC      molecules are also useful as probes for gene mapping, or for identifying
CC      E. faecalis in biological samples. The kit and methods are useful for
CC      detecting Enterococcus antibodies or nucleic acid molecules in a
CC      biological sample. The present sequence is a novel E. faecalis nucleic
CC      acid of the invention. Note: The sequence data for this patent did not
CC      form part of the printed specification, but was obtained in electronic
CC      format directly from the USPTO at
CC      seqdata.uspto.gov/sequence.html?docID=20030017495
XX
SQ      Sequence 1687 BP; 632 A; 284 C; 360 G; 411 T; 0 U; 0 Other;

Query Match
3.3%; Score 49.8; DB 7; Length 1687;
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 22:18:49 ; Search time 4180 Seconds  
(without alignments)  
10780.390 Million cell updates/sec

Title: US-10-009-254-1  
Perfect score: 1509  
Sequence: 1 atgaaagaagaatgattca.....gtcgtcgttaccgttcttaa 1509

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Database :

EST:

1: em\_estba:\*

2: em\_estba:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hcc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	76	5.0	843	28	B2643413
2	69.8	4.6	844	29	BX139987
3	69.4	4.6	494	29	FR0048073
C 4	69	4.6	829	29	BX173672
-----					
C 1	76	5.0	843	28	B2643413
2	69.8	4.6	844	29	BX139987
3	69.4	4.6	494	29	FR0048073
C 4	69	4.6	829	29	BX173672

5	68	4.5	989	29	CNS02HA4
C 6	67.8	4.5	641	28	AQ946120
7	67	4.4	619	29	CG372306
8	66.4	4.4	501	29	FR0048173
9	66.4	4.4	699	13	BX729701
10	66.2	4.4	575	28	BZ780846
C 11	66.2	4.4	1101	29	CNS012TP
C 12	65.2	4.3	1101	29	CNS0039G
C 13	64.6	4.3	735	29	CNS04NSM
14	62.8	4.2	480	28	BZ643398
15	62.6	4.1	513	28	BZ423698
C 16	61.8	4.1	994	13	BX414650
C 17	61.4	4.1	470	29	FR0018463
18	61	4.0	450	29	FR0025683
C 19	60.8	4.0	479	28	BH211305
20	60.6	4.0	788	28	AZ183942
21	60	4.0	525	28	BZ351620
C 22	60	4.0	806	29	CNS04AEE
23	59.8	4.0	773	29	CNS01VTG
24	59.6	3.9	617	12	BJ391130
C 25	58.6	3.9	427	28	BZ422321
26	58.6	3.9	707	29	CG403581
27	58.4	3.9	914	28	AZ205202
C 28	58	3.8	548	28	BH200391
29	58	3.8	619	29	FR0006944
30	57.6	3.8	641	29	CE682324
31	57.4	3.8	619	29	FR0047601
C 32	57.4	3.8	781	29	BH145762
33	56.8	3.8	587	28	BH873429
C 34	56.8	3.8	856	28	BH178274
C 35	56.8	3.8	856	29	CNS07KML
36	56.8	3.8	1200	13	BX437758
C 37	56.6	3.8	713	12	BJ408597
C 38	56	3.7	541	12	BJ406947
39	56	3.7	854	28	AQ411842
40	55.8	3.7	363	29	CG807311
C 41	55.8	3.7	670	28	BZ388480
42	55.6	3.7	471	29	CG807194
43	55.4	3.7	933	28	AZ204694
44	55.2	3.7	350	28	BH879665
C 45	55.2	3.7	655	12	BJ403823
C 46	55.2	3.7	657	12	BJ406829
47	55.2	3.7	694	28	BH355163
48	55	3.6	550	29	FR0043207
C 49	55	3.6	592	12	BJ409049
C 50	55	3.6	732	12	BJ438152
C 51	54.8	3.6	646	12	BJ408562
C 52	54.8	3.6	661	12	BJ408664
53	54.8	3.6	700	28	AQ940248
54	54.6	3.6	356	28	CC171914
C 55	54.6	3.6	557	12	BJ403416
C 56	54.6	3.6	593	12	BJ403193
C 57	54.6	3.6	696	12	BJ380445
C 58	54.6	3.6	706	12	BJ406949
C 59	54.6	3.6	785	12	BJ34750
C 60	54.6	3.6	857	12	BJ403229
61	54.6	3.6	989	28	CC149236
62	54.4	3.6	318	9	AU086474
63	54.4	3.6	395	28	BZ316940
C 64	54.2	3.6	463	12	BJ390675
C 65	54.2	3.6	530	28	AZ166409
66	54.2	3.6	544	12	BG226227
C 67	54.2	3.6	664	12	BJ408060
C 68	54.2	3.6	714	12	BJ406644
C 69	54	3.6	614	12	BJ405374
C 70	54	3.6	702	12	BJ407918
71	53.8	3.6	306	28	BZ375744
72	53.8	3.6	334	28	BZ614791
C 73	53.8	3.6	602	29	DR7024T
74	53.8	3.6	619	29	FR0047503
C 75	53.8	3.6	796	28	BH180509
C 76	53.8	3.6	796	29	CNS07MCJ
77	53.6	3.6	500	9	AU088119





```
Query Match      4.6%; Score 69.8; DB 29; Length 844;
Best Local Similarity 46.3%; Pred. No. 1.8e-05;
Matches 230; Conservative 0; Mismatches 267; Indels 0; Gaps 0;

QY 137 ATAAAGCATATAAAGTTTTTGTATGACAGAAATAGATAATCGAATGATCTGATTCGAATA 196
    |||||
Db 348 ATATAATAATAATATGATAATAATAATGATAATAATAATGATAATAATAATAATAATA 407

QY 197 AAGATGGAGCTTCTTTATTTTAAATTCCTCAAGGTAAGAAGCTGAGTATATAAGCTTCAACTG 256
    |||||
Db 408 TTGATAATAATAATGATAATAATAATAATGATAATAATAATAATAATAATAATAATAATA 467

QY 257 ATTTTAAATCTCTTTTACGACAACTACTAATGAGGGGAGAACATATGTAACATAAAAAAG 316
    |||||
Db 468 ATAATAATAATAATGATAATAATAATAATATTTGATAATAATTTGATGATAATAATAATAATG 527

QY 317 ATACTGGCTGCAGCAAAATGAGTTGGACATGGCTGGCTAAATCTATATCAGCTATATCTACAC 376
    |||||
Db 528 ATATAATAATAATAATGATAATAATAATAATGATAATAATAATAATAATAATAATAATAATA 587

QY 377 CAGTTTCCACTGTTACTGAGTCAAAATGATGGTACTGAGGTTATTAATGTTTCCCAAT 436
    |||||
Db 588 ATGATGATAATAATAATGATGATAATAATGATAATAATAACAAATAATAATAATAATGATA 647

QY 437 ATGGATATTTATGTTTCTAGCAGCTGTTAATAATGAGAGCTGTAATATGTTGTTACATCTG 496
    |||||
Db 648 ATAATAATAATAATGATGATAATAATAATAATAATAATGATGATAATAATAATAATAATGATA 707

QY 497 TAACTCCAAATCTACTATTTTCATGAAAGAACTACTGATCGCATGGGGAGAGTGGTG 556
    |||||
Db 708 CTAATAATGACAATAATAATAATAATGACAATAATAATAATAATAATAATAATAATAATGATA 767

QY 557 GAAAACTGTAGATCAAAAAAGCTACTCGTTGGTGATACAGTCAAAATATATCTATTACTT 616
    |||||
Db 768 ATAATAATGACTACTATAATAATGATAATAATAATAATAATAATAATAATAATAATAATGATA 827

QY 617 ATAAGAATGCAGTCAAT 633
    |||||
Db 828 ATAATAATGATGATAAT 844

RESULT 3
FR0048073
LOCUS      494 bp      DNA      linear      GSS 05-JAN-2001
DEFINITION Fugu rubripes GSS sequence, clone 263K15BD8, genomic survey
ACCESSION  AL444858.1  GI:12052694
VERSION     GSS; genome survey sequence.
KEYWORDS    Takifugu rubripes (Fugu rubripes)
SOURCE      Takifugu rubripes
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE   1 (bases 1 to 494)
            Direct Submission
            Author(s) Clark, M.S.
            Title     Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
            Journal   Centre Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmrc.ac.uk
            Vector: pBluescript II KS
            V type: phagemid
            PRIMER: KS
            DESC:     One pass dye-terminator sequencing of BAC (pBelobACII) cloned.
            genomic sequence
            The BACs can be obtained from http://www.incyte.com.
FEATURES    source
            Location/Qualifiers
                1..494
                /organism="Takifugu rubripes"
                /mol_type="genomic DNA"

Query Match      4.6%; Score 69; DB 29; Length 829;
Best Local Similarity 45.2%; Pred. No. 2.7e-05;
Matches 252; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

/db_xref="taxon:31033"
/clone="263K15BD8"
/clone_lib="BAC 263K15"

ORIGIN
Query Match      4.6%; Score 69.4; DB 29; Length 494;
Best Local Similarity 48.6%; Pred. No. 2.1e-05;
Matches 190; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 236 CTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTACTAATGAGGGA 295
    |||||
Db 22 CTGGGCTTCTACTACTACTAGTACTGCTGCTACTACTACGACTACGACTACTGCTGCTGCT 81

QY 296 GAACATATGTAACATAAAGATATCTGCGTCAGCAATGAGATGCGCATGGGCTAAAT 355
    |||||
Db 82 CGACTACTGCGACGAGCTGCTACTACTGCTGCGGCTGCTTCTACTACTGCTACTACTA 141

QY 356 CTATATCAGCTAATATACACAGTTCCTACTGTTTACTGAGTCAATATGATGCTACTG 415
    |||||
Db 142 CGAGACTGCTACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTA 201

QY 416 AGTTTAAATGTTTCCCAATATGATATTTATGTTTCTAGCACTGTTAATAATGAG 475
    |||||
Db 202 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 261

QY 476 CTGTAATTTATGTTTACTACTCTGTAATCTCAATGCTACTATTTCATGAAAGAAATCTGATG 535
    |||||
Db 262 CTACTACTACTACTACTACTACTACTGCTGCTACTGCTACTACTACTACTACTACTACTACTA 321

QY 536 CGACATGGGAGAGTGGTGGGAAAACTGTAGATCAAAAAAGCTACTCGTTGGTGATA 595
    |||||
Db 322 CTACTACTACTACTACTGCTGCTGCTGCTGCTGCTACTACTACTACTACTACTACTACTGCTA 381

QY 596 CAGTCAAAATATATCTATTCTTATAAGATGC 626
    |||||
Db 382 CTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGCTA 412

RESULT 4
BX173672/c
LOCUS      829 bp      DNA      linear      GSS 13-MAR-2003
DEFINITION Danio rerio genomic clone DKEY-150M6, genomic survey sequence.
ACCESSION  BX173672.1  GI:28005377
VERSION     GSS.
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 829)
            Humphray, S.J., Huckle, E. and Durham, J.L.
            Direct Submission
            Title     Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Journal   Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Unpublished
            This sequence was generated from the SP6 end of BAC 150M6. 150M6 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES    source
            Location/Qualifiers
                1..829
                /organism="Danio rerio"
                /mol_type="genomic DNA"
                /db_xref="taxon:7955"
                /clone="DKEY-150M6"
                /tissue_type="testis"
                /note="vector pIndigoBAC-536"

ORIGIN
Query Match      4.6%; Score 69; DB 29; Length 829;
Best Local Similarity 45.2%; Pred. No. 2.7e-05;
Matches 252; Conservative 0; Mismatches 305; Indels 0; Gaps 0;
```

```
QY 385 ACTGTTACTAGTCAAAATATGATGCTACTAGGTTTAAATGTTTCCCAATATGATAT 444
Db 645 AATAAATAAATAAGATAATAATAATAATGATGATGATGATGATGATGATGATGATGAT 586
QY 445 TATTATGTTTCTAGCACTGTTAATAATGAGCTGTAATTTATGTTTACATCTGTAATCCCA 504
Db 585 AATAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 526
QY 505 AATGCTACTATTATCAAGAAAGATATCTGATGCGACATGGGAGATGGTGGTGGAAAAACT 564
Db 525 AATAAATAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 466
QY 565 GTAGATCAAAAACGCTACTCGGTTGCTGATACAGTCAAAATATATTTACTTTAAGAT 624
Db 465 GATGATGATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 406
QY 625 GCAGTCAAAATATCATCGTACAGAAAAAGTGTATCAATATGTTAATAAGGATATCTATGCCA 684
Db 405 GATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 346
QY 685 TCTGCTTCTGATGTTGATGACGAAGGCTCTTATGAAGTAACTATTTACTGATGATCA 744
Db 345 AATAAATAAATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 286
QY 745 GGAATAATTTACAACTCTAACTCAAGCTTCGGAAGAAAGCAACTGGGAAGTATAACCTGTTA 804
Db 285 AATAATAGAAATAATAATAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 226
QY 805 GAGGAAAAATAAATTTTCAGATTACTATTTCCGTTGGGCGACTACCAATACTCCAAACCGGA 864
Db 225 AATAATAATAATAATAAGAAATAATAATAATGATGATGATGATGATGATGATGATGATGAT 166
QY 865 AATACTCAAAATCGAGCTAATGATGACTTTTATTAAGGGAATAATAACAATCACAGTC 924
Db 165 AATAGTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 106
QY 925 ACTTATACAGGAGTATT 941
Db 105 TAAATAGAAATAAT 89

RESULT 5
CNS02HA4
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
139021 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL197365
VERSION
GI:7835515
GSS; genome survey sequence.
Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau L., Fisher,C.,
Bernot,A., Fzames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
JOURNAL
FUBMED
10835645
2
Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
Fzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

MEDLINE
PUBMED
10899143
REFERENCE
3 (bases 1 to 989)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
COMMENT
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
source
1..989
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="1139021"
/clone_lib="G"
/note="Genoscope sequence ID : COAG139AH11LPL1-end : T7"
ORIGIN
Query Match 4.5%; Score 68; DB 29; Length 989;
Best Local Similarity 44.9%; Pred. No. 4.6e-05;
Matches 213; Conservative 7; Mismatches 254; Indels 0; Gaps 0;
QY 171 TAATCAAAATGATCTGATTCGAATAAAGATGGAGCTTCTTTATTTAAATTCCTCAAGTAA 230
Db 56 TACYACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 115
QY 231 AGAAGCTGAGTATAAGCTTCAACTGATTTTAAATTTCTCTTTTACGACAATCTAATG 290
Db 116 TAATACAAAACACTACWACWACTACTACTACTACTACTACTACTACTACTACTACTACTAA 175
QY 291 AGGAGACATATGTAACTAAAAGATGCTCGTCAGCAATAGATGGGACATGGC 350
Db 176 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 235
QY 351 TAAATCTATATCAGCTAATACTACACCACTTTCCACTCTTACTAGTCAAAATATGATGG 410
Db 236 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 295
QY 411 TACTGAGGTTAATGTTTCCCAATATGGATATTTATTTATGTTTCTAGCACCTGTTAATAA 470
Db 296 TACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 355
QY 471 TGGAGCTGTAATTTGTTTACATCTGTAATCTCAATGCTACTATTTCATGAAAAGATAC 530
Db 356 TACTGCTAATACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 415
QY 531 TGATCGACATGGGAGATGGTGGGAAAACCTAGATCAAAAACGTAATCGTGGTGG 590
Db 416 TACYACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTGG 475
QY 591 TGATCAGTCAATATATCTACTTACTTAAAGATCGAGTCAATATCATGTTAC 644
Db 476 TGGTCTCTCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTCC 529

RESULT 6
AQ946120/c
LOCUS
DEFINITION
Sheared DNA-46J23-TR Sheared DNA Trypanosoma brucei genomic clone
Sheared DNA-46J23, genomic survey sequence.
AQ946120
ACCESSION
VERSION
AQ946120.1 GI:6769385
KEYWORDS
GSS.
SOURCE
Trypanosoma brucei
ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 641)
REFERENCE
El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
```

Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S.,  
 Donelson, J., Fraser, C. and Adams, M.  
 Determination of clone end sequences from Trypanosoma brucei GUTat  
 10.1 sheared DNA library  
 Unpublished (1999)  
 Other GSSs: Sheared DNA-46J23.TP  
 Contact: Najib M. El-Sayed  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: nelsayed@tigr.org  
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
 DNA library constructed at TIGR. Clones will be available for  
 distribution through ATCC. Sheared DNA end sequences search page:  
<http://www.tigr.org/tldb/mdb/tbdb/>.  
 Seq primer: M13-Reverse  
 Class: shotgun.

# FEATURES

source  
 1. .641  
 /organism="Trypanosoma brucei"  
 /mol\_type="genomic DNA"  
 /strain="TREU927/4 GUTat 10.1"  
 /db\_xref="taxon:5691"  
 /clone\_lib="Sheared DNA-46J23"  
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from a cloned population of  
 Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically  
 sheared to give a tight size distribution (approx 2 kb).  
 The v + i method used for the library construction is  
 described in detail in Smith, H.O. and Venter, J.C.  
 (Making small insert libraries for whole genome shotgun  
 sequencing projects. In Genome Sequencing: A Practical  
 Approach, eds. M. Vaudin and B. Borell, Oxford University  
 Press, 1999)."

# ORIGIN

Query Match 4.5%; Score 67.8; DB 28; Length 641;  
 Best Local Similarity 45.3%; Pred. No. 5e-05;  
 Matches 246; Conservative 0; Mismatches 297; Indels 0; Gaps 0;  
 QY 366 TAATACTACACAGTTTCCACTGTTTACTGAGTCAAAATATGATGCTACTGAGGTTATTAA 425  
 Db |||||  
 QY 621 TAATAGTACTACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 562  
 Db |||||  
 QY 426 TGTTCCTCCAAATGATATATTATGTTCTTAGCTGTTTAAATGAGGCTGTAATTAT 485  
 Db |||||  
 QY 561 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 502  
 Db |||||  
 QY 486 GGTTCATCTGTAACCTCCAAATGCTACTATTTCATGAAAGAACTACTGATGCGACATGGG 545  
 Db |||||  
 QY 501 TAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 442  
 Db |||||  
 QY 546 AGATGGTGGTGGAAAACCTGTAGATCAAAAACGCTACTCGGTGTGTGATACAGTCAAAATA 605  
 Db |||||  
 QY 441 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 382  
 Db |||||  
 QY 606 TACTATTACTTATAGAGATGAGTCAATATCATCGGTACAGAAAAGTGTATCAATATGT 665  
 Db |||||  
 QY 381 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 322  
 Db |||||  
 QY 666 TATTAAGGATACCTATGCCATCTGCTTCTGTAGTCTTGTGACGAGGCTCTTATGAGT 725  
 Db |||||  
 QY 321 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 262  
 Db |||||  
 QY 726 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 785  
 Db |||||  
 QY 261 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 202  
 Db |||||  
 QY 786 TGGGAAGTATACCTGTTAGAGGAAAATAATAATTTTCACGATTACTATTCCTGGGCGAG 845  
 Db |||||

Db 201 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 142  
 QY 846 TACCAATACTCCACCGGAAATACTCAAAATGGAGCTAATGATGACTTTTTTTATAAGG 905  
 Db |||||  
 Db 141 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 82  
 QY 906 AAT 908  
 Db 81 AGT 79

# RESULT 7

CG372306  
 LOCUS  
 DEFINITION  
 OGA8108TC ZM 0.7 1.5 KB Zea mays genomic clone ZMMBma0811B15,  
 genomic survey sequence.  
 CG372306  
 ACCESSION  
 CG372306.1 GI:34289573  
 VERSION  
 GSS.  
 KEYWORDS  
 Zea mays  
 SOURCE  
 ORGANISM  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 REFERENCE  
 1 (bases 1 to 619)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 CONTACT: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.

# FEATURES

source  
 1. 619  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBma0811B15"  
 /clone\_lib="ZM 0.7 1.5 KB"  
 /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

# ORIGIN

Query Match 4.4%; Score 67; DB 29; Length 619;  
 Best Local Similarity 45.5%; Pred. No. 7.6e-05;  
 Matches 238; Conservative 0; Mismatches 285; Indels 0; Gaps 0;  
 QY 48 TGGTATGGCTGTATCACCAGTTACGCCGATAGCTTTTCCCGCTCAGACAGGACAATTAC 107  
 Db |||||  
 QY 53 TGCTACTGCTGCTACTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTAC 112  
 Db |||||  
 QY 108 AGTTCAAGATACCTCAAAAGCGCAACCTATAAAGCATATAAAGTTTGTGATGAGAAT 167  
 Db |||||  
 QY 113 TGCTACTACTACTACTACTACTACTCTCTACTACTACTACTACTACTACTACTACTAC 172  
 Db |||||  
 QY 168 AGATAATGCAATGCTATCTGATTCGAATAAAGATGGAGCTTCTTTATTTTAAATTCCTCAAGG 227  
 Db |||||  
 QY 173 TACTACTACTACTACTACTACTACTACTACTACTACTGCTGCTGCTGCTACTACTGCT 232  
 Db |||||  
 QY 228 TAAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTACGACAACCTACTAA 287  
 Db |||||  
 QY 233 TACTACTACTGCTACAACCTACTCTGCTACTACTACTGCTACTACTACTGCTACTACTAC 292  
 Db |||||  
 QY 288 TGGAGGAGGAACATATGTACTAAAAAAGATACCTGCGTCAGCAATAGATTTGGACATG 347  
 Db |||||

D	b		293	TGCTACTACTACGCCCACTACTGCTACTACTCGTCTACTACTGCTACTACTGCTACTACTGC	352
Q	y		348	GGCTAAATCTATACAGTAATACTACCAGTTTCACGTGTTTCCAGTGTTACTCAGTCAAATAATGA	407
D	b		353	TACTACTGCTACTGCTACTACTACTACTACTGCTACTGCTACTACTGCTACTACTGCTAC	412
Q	y		408	TGTTACTGAGGTTATTAAATGTTTTCCCNAATATGGANATATATATGTTTCTTAGCACTGHTAA	467
D	b		413	TACTGCTACTGCTCCTACTGCTGCTACTGCTGCTACTGCTGCTACTGCTACTACTACTGCTAC	472
Q	y		468	TAATGGAGCTGTAATATATGTTTACATCTGAACCTCCAATAGCTACTATTTCATGAAAAGAA	527
D	b		473	TACTTCTACTACTACTACTACTGCTACTGCTACTACTGCTACTACTGCTACTGCTATTGCTGCTAC	532
Q	y		528	TACTGATGCCACATGGCGAGATGGTGGTGAAAAAAGTATAGAT	570
D	b		533	TGCTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTGCT	575
RESULT 8					
FR0048173			501 bp	DNA	linear GSS 05-JAN-2001
LOCUS			Fugu rubripes GSS sequence, clone 264E22cA9,		genomic survey
DEFINITION			sequence.		
ACCESSION			AL444958		
VERSION			AL444958.1 GI:120533458		
KEYWORDS			GSS; genome survey sequence.		
SOURCE			Takifugu rubripes (Fugu rubripes)		
ORGANISM			Takifugu rubripes		
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
			Actinopterygii; Neopterygii; Teleostei; Euteleoestei; Neoteleosteoi;		
			Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;		
			Tetraodontoidea; Tetraodontidae; Takifugu.		
			1. (bases 1 to 501)		
REFERENCE			Clark,M.S.		
AUTHORS			Direct Submission		
TITLE			Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource		
JOURNAL			Centre Hinxton, Cambridge, CB10 1SB. UK Email:		
			biohelp@hmp.mrc.ac.uk		
COMMENT			Vector: pBluescript II KS		
			V type: phagemid		
			PRIMER: KS		
			DBSCR:		
			One pass dye-terminator sequencing of BAC (pBelobACII) cloned		
			genomic sequence		
			The BACs can be obtained from http://www.incyte.com.		
FEATURES			Location/Qualifiers		
			source		
			1..501		
			/organism="Takifugu rubripes"		
			/mol_type="genomic DNA"		
			/db_xref="taxon:31033"		
			/clone="264E22cA9"		
			/clone_lib="BAC 264E22"		
ORIGIN					
Query Match			4.4%; Score 66.4;	DB 29;	Length 501;
Best Local Similarity			48.9%;	Pred. No. 0.0001;	
Matches 178;			Conservative	Mismatches 186;	Indels 0; Gaps 0
Q	y		171	TAATGCAATGTTATCTGATTCGAATAAAGATGGAGCTTCTTATTATTAATTCCTCAAGGTAA	230
D	b		101	TACTACTACTGCTACTACTACTACTACTACTACTACTGCTGCTCTTCTACTGCTACTAC	160
Q	y		231	AGAAGCTGAGTATAAAGCTTCAACGTGATTTTAATTTCTCTTTTACGACAAC TACTAATGG	290
D	b		161	TACTACTACTGCTACTACTACTGCTGCTCTTCTACTACTACTACTACTACTACTACTACTAC	220
Q	y		291	AGGGAGAACATATGTAAC TAAAAAAGATAGCTGCGTCAGCAATGAGATTGGACATGGCG	350
D	b		221	TACTACTACTGCTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTCTGCTG	280
Q	y		351	TAAATCTATATACGCTAATATPACACCAAGTTTCCACTGTTTACTGAGTCAATATATGATGG	410

[illegible]

RESULT 10					
BZ780846					
LOCUS	BZ780846	575 bp	DNA	linear	GSS 14-MAR-2003
DEFINITION	ii20h11.g1 WGS-ShicolorF (DH5a methyl filtered) Sorghum bicolor genomic clone ii20h11, genomic survey sequence.				

02730640.1 G1.28530586  
 GSS.  
 Sorghum bicolor (sorghum)  
 Sorghum bicolor  
 Sorghum bicolor  
 ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 575)  
 Rabinowicz,P.D., O'Shaughnessy,A.L., Balijs,V., Dedhia,N.,  
 Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,  
 Zutavern,T., Palmer,L., McComb,W.R. and Martienssen,R.A.  
 Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)  
 Unpublished (2002)  
 TITLE  
 JOURNAL

CONTACT: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: ii20 row: h column: 11  
Seq primer: -21M13Univrev  
Class: shotgun  
High quality sequence stop: 575.  
Location/Qualifiers  
source 1..575

in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using Sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (.x/y reads in M13mp19 .b/g reads in pUC19). The same ligation was transformed into DH5a."

## ORIGIN

[illegible]

RESULT 11  
CNS012TP/c  
LOCUS  
DEFINITION  
CNS012TP 1101 bp DNA linear GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN08A01 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION  
VERSION  
AL102007  
AL102007.1 GI:5613618  
KEYWORDS  
GSS.  
SOURCE  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage  
JOURNAL  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr)  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))





Contact: Cathy Whitelaw  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

## FEATURES

Location/Qualifiers

1. .480

source

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA011J17"

/clone\_lib="ZM\_0.7\_1.5\_KB"

/notes="Vector: pBCSK-; Site\_1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

## ORIGIN

Query Match 4.2%; Score 62.8; DB 28; Length 480;  
Best Local Similarity 48.8%; Pred. No. 0.0007;  
Matches 200; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 130 GCAACCTATAAGCATATAAGTTTTTGATGCAGAAATAGATAATGCAAAATGATCTGAT 189  
DB 17 GCTACTACTGCTACTACTGCTACTGCTGCTGCACTGCTACTGCTACTGCTACTGCT 76

QY 190 TCGAATAAGATGAGCTTCTTATTAATTCCTCAAG---GTAAAGAGCTGAGTATAA 246  
DB 77 ACTACTACTACTGCTGCTACTACTACTACTGCTGCTGCTACTACTACTACTGCTACT 136

QY 247 GCTTCAACTGATTTAATTCCTTTTACGACAACTACTAAATGAGGGAGAACATATGA 306  
DB 137 ACTACTACTACTGCTGCTACTACTGCTACTGCTGCTGCTGCTGCTGCTACTACTGCT 196

QY 307 ACTAAAAAGATAGCTCGTCAGCAAAATGAGATTGCGACATGGCTAAATCTATACAGCT 366  
DB 197 ACTGCTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256

QY 367 AATCTACACCACTTTCCAGTGTACTGAGTCAAAATGATGGTACTGAGGTATTATAT 426  
DB 257 ACTACTGCTACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACT 316

QY 427 GTTTCCTCAATGATATTAATGTTTCTAGCACTGTTAAATGAGCTGTAATATG 486  
DB 317 GCTACTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 376

QY 487 GTTACATCTGTAATCCAAATGCTACTATTTCATGAAAGAAATGATGTC 536  
DB 377 ACTGCTGCTACTGCTACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 426

## RESULT 15

BZ423698

LOCUS

DEFINITION BZ423698 513 bp DNA linear GSS 10-DEC-2002  
id52c10.g1 WGS-SbicolorF (DH5a methyl filtered) Sorghum bicolor  
genomic clone id52c10 5', genomic survey sequence.

ACCESSION BZ423698

VERSION BZ423698.1

KEYWORDS GSS.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 513)

AUTHORS Rabinowitz,P.D., O'Shaughnessy,A.L., Ballja,V., Dedhia,N.,

Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,

Zutavern,T., Palmer,L., McCombie,W.R. and Martienssen,R.A.

TITLE Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)

JOURNAL Unpublished (2002)

COMMENT Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mccombie@cshl.org  
Plate: id52 row: c column: 10  
Seq primer: -21M13UnivRev  
Class: shotgun  
High quality sequence stop: 513.

## FEATURES

source

1. .513

/organism="Sorghum bicolor"

/mol\_type="genomic DNA"

/db\_xref="taxon:4558"

/clone="id52c10"

/lab\_host="DH5a"

/clone\_lib="WGS-SbicolorF (DH5a methyl filtered)"

/note="Site 1: Xba I; Site 2: Xba I; The vector was  
digested with XbaI and one nucleotide was added by fill in  
in the recessive 3' end. The genomic DNA was nebulized,  
end repaired, adaptor ligated and size fractionated using  
sephadex. The resulting fragments were between 0.8 and 3  
kb and were cloned into the vector (.X/Y reads in M13mp19,  
.b/g reads in pUC19). The same ligation was transformed  
into DH5a."

## ORIGIN

Query Match 4.1%; Score 62.6; DB 28; Length 513;  
Best Local Similarity 46.3%; Pred. No. 0.00078;  
Matches 206; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

QY 171 TAATGCAATGATGATTCGATTCGAATAAGATGGAGCTTCTTATTTAAATTCCTCAAGGTAA 230  
DB 65 TACT 124

QY 231 AGAAGCTGAGTATAAGCTTCAACTGATTTTAAATTCCTTTTACGACAACTACTAAATGG 290  
DB 125 TACT 184

QY 291 AGGGAGACATATGTAACCTAAAGAGATCTGCGTCAGCAATGAGATGCGACATGGCG 350  
DB 185 TAGTGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 244

QY 351 TAAATCTATATCAGCTAATATCTATACACCACTTTCCTACTGTCTAGTCAAAATGATGG 410  
DB 245 TACT 304

QY 411 TACTGAGGTTAATATGTTTCCCAATATGGATATTATATTGTTCTAGCACCTGTTAATAA 470  
DB 305 TACT 364

QY 471 TGGAGCTGTAATATGTTTACATCTGTAACCTCAATGCTACTATTTCATGAAAAGAAATAC 530  
DB 365 TACT 424

QY 531 TGATGCGACATGGGAGATGTTGGTGGAAAACTGTAGATCAAAAAACGTACTCGGTGG 590  
DB 425 TGCTGCTGCTACTACTACTACTGCTGCTGCTACTACTACTGCTGCTACTACTACTACTGC 484

QY 591 TGATCAGTCAAAATATATCTATTACT 615  
DB 485 TGCTACTACTACTACTACTACTACTACT 509

## RESULT 16

BZ414650/c

LOCUS

DEFINITION BZ414650 Homo sapiens THYMUS Homo sapiens cdna clone CS0CAP001YN02

3-PRIME, mRNA sequence.

ACCESSION BZ414650

VERSION BZ414650.1

KEYWORDS GI:30763455

EST.

BX414650 994 bp mRNA linear EST 15-MAY-2003  
BX414650 Homo sapiens THYMUS Homo sapiens cdna clone CS0CAP001YN02  
3-PRIME, mRNA sequence.  
BX414650 BX414650  
BX414650.1 GI:30763455  
EST.



SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 994) Li,W.B., Gruber,C., Jessee,J. and Polayes,D. Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope
AUTHORS	Genoscope - Centre National de Sequencage
TITLE	BP 191 91006 EVRY cedex - France
JOURNAL	Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
COMMENT	Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6015.f Contact: Feng Liang Email : fliang@lifetech.com URL : <a href="http://fulllength.invitrogen.com/">http://fulllength.invitrogen.com/</a> Invitrogen Corporation 1600 Faraday Avenue Genosope sequence ID : CSOCAP001DG0INP1. Location/Qualifiers
FEATURES	1..994 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /cclone="CSOCAP001YN02" /tissue_type="THYMUS" (clone_lib="Homo sapiens THYMUS" note=vector: pcMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pcMVSPORT 6 vector. Library was not normalized."
ORIGIN	Query Match                 4.1%; Score 61.8; DB 13; Length 994; Best Local Similarity    22.7%; Pred No. 0.0012; Matches 138; Conservative 197; Mismatches 255; Indels 8; Gaps 1;
Qy	330 AAATGAGATTGGCGACATGGGTAAACTATATCATCAGCTAATACTACACCAGTTTCCTCACTGT 389
Dd	968 AARDRWAGAARARRADWDWDAAWAAWRDWTAWWAARKKWKDWTTTTTKTKKDADA 909
Qy	390 TACTGAGTCAATTAATCATGGTGACTGAGGTATTAAATGTTTCCCAATATCGATATATTATTA 449
Dd	908 DWGDWDWDWDTAWWWAADDDWAAARWAAAAATWWWTDWTAMWTRDKATWKAADDADDDAKTK 849
Qy	450 TGTTTCTTAGCACTGTTAATGAATGAGCTGTAATATTATGTTTACATCTCTGAATCTCAAATGC 509
Dd	848 TTITTTTWKAWTDWAGTAKWNADATWADADAWARAARWAAARWRAAAAANWWAADAD 789
Qy	510 TACTATTTCATGAAAGAATACTGATCGCATCGGGAGATGGTGGTGGA AAAAACTGTAGA 569
Dd	788 WWWWWDAADAAMWAAWDTKDKRARAADWARAAAAAAWWWGRRGRADAAGDKDKARAA 729
Qy	570 TCASAAAACGTACTCTGGTGGTGATACAGTCAATATATCTATTACTTATTAAGATGCGAGT 629
Dd	728 WWDGAGGRDXDRDWAERWDDADAAARBARAAAAATAWWATWKKDKDAAAA----- 676
Qy	630 CAATTATCTAGTGTACAGAAAAAGTGTTATCAATATGTTATAAGGATACTATGCCATCTGC 689
Dd	675 -AWDTAGKGKDDARDWWAAWTTTTTTTTTTTAAWAADKAKWKTATTTTTTTTTTTTTTT 617
Qy	690 TTCTGTAGTTGATTTGAACAAGGGTCTTATGAAGTAACTATTACTGATGGATCAGGGAA 749
Dd	616 TTTTKTKTTAKWTDWADWDWTTTTTTTTTTTTTTTTTTTTTAAATTTKTKTKTKTKTKTTT 557
Qy	750 TAITTACAACCTTA ACTCAAGTTTCGGAAAAAGCAACTGGGAAGTATAAACCTGTGTAGGA 809
Dd	556 TRDVTDAARARRRR RRRRRRRRRRGRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRAA 497
Qy	810 AAATAATAATTTTACAGTATTACTATTTCCGTGGGAGCTACCAATACTCCAACCGGAATAC 869
Dd	496 ARRAANNRRDRNRKNNTKTKWAWTWATRKRGRTKTGTGTGTGTGTGTGTGTGTGTGTGT 437
Qy	870 TCAAAATGGAGCTAATGATGACTTTTTTTTTTATAAGGGAATAAATACATCACAGTCACTTA 929

LOCUS PR0025683 450 bp DNA linear GSS 10-DEC-1997  
DEFINITION F.rubripes GSS sequence, clone 154E17aC12, genomic survey sequence.  
ACCESSION AL018519  
VERSION AL018519.1 GI:2684887  
KEYWORDS GSS; genome survey sequence.  
SOURCE Takifugu rubripes (Fugu rubripes)  
ORGANISM Takifugu rubripes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Takifugu.  
REFERENCE 1 (bases 1 to 450)  
AUTHORS Elgar, G., Clark, M., Smith, S., Meek, S., Warner, S., Umrana, Y.,  
Williams, G. and Brenner, S.  
TITLE Direct Submission  
JOURNAL Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource  
Centre Hinxton, Cambridge, CB10 1BS. Email: biohelp@hmp.mrc.ac.uk  
COMMENT Vector: pBluescript II KS  
V type: phagemid  
PRIMER: KS  
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic  
sequence.  
FEATURES  
source  
1..450  
/organism="Takifugu rubripes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:31033"  
/clone\_lib="154E17aC12"  
/clone\_lib="cosmid 154E17"  
ORIGIN  
Query Match 4.0%; Score 61; DB 29; Length 450;  
Best Local Similarity 49.7%; Pred. No. 0.0018;  
Matches 151; Conservative 0; Mismatches 153; Indels 0; Gaps 0;  
Qy 236 CTGAGTATAAGTTCACAGTATTAAATCTCTTTTACGACAACTACTAATGGAGGA 295  
Db |||||  
Qy 138 CTGCTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 197  
Db |||||  
Qy 296 GACATATGTAACATAAAGATCTGCGTCAGCAATGAGATGGACATGGGCTAAAT 355  
Db |||||  
Qy 198 CTGCTGCTGCTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACT 257  
Db |||||  
Qy 356 CTATATCAGCTAATACACAGTTTCCACTGTTACTGAGTCAAAATATGATGCTACTG 415  
Db |||||  
Qy 258 CTACTGCTACTACTGCTACTGCTGCTACTGCTACTGCTACTGCTACTGCTACTGCT 317  
Db |||||  
Qy 416 AGGTATTAATGTTTCCAAATGGAATTAATTAATGTTTCTAGCAGTGTAAATGAG 475  
Db |||||  
Qy 318 CTATTAATGTTTCTACTGCTGCTGCTACTACTGCTACTACTACTACTGCTGCTGCT 377  
Db |||||  
Qy 476 CTGTAATTAATGTTTCTGTAATCTCAATGCTACTATTCATGAAGAATGATGATG 535  
Db |||||  
Qy 378 CTGCTACTGCTACTACTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTG 437  
Db |||||  
Qy 536 CGAC 539  
Db |||||  
Qy 438 CGAC 441  
Db |||||  
RESULT 19  
BH211305/c  
LOCUS BH211305  
DEFINITION Sm1-50G22.TF Sm1 Schistosoma mansoni genomic clone Sm1-50G22,  
genomic survey sequence.  
ACCESSION BH211305  
VERSION BH211305.1 GI:16391571  
KEYWORDS GSS.  
SOURCE Schistosoma mansoni  
ORGANISM Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;  
Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.

REFERENCE 1 (bases 1 to 479)  
AUTHORS Shetty, J., Simpson, A., Malek, J., Koo, H., LoVerde, P.T. and  
El-Sayed, N.M.  
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)  
JOURNAL Sm1 BAC library for gene discovery and map construction  
COMMENT Unpublished (2001)  
Other GSSs: Sm1-50G22.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Schistosoma mansoni (Puerto Rico) Sm1  
BAC library. For clone availability, please contact Dr. Najib  
El-Sayed at tigr (nelsayed@tigr.org) or Dr. Phillip LoVerde at  
State University of New York, Buffalo, New York, USA  
(loverde@buffalo.edu)  
Seq primer: M13 For  
Class: BAC ends.  
FEATURES  
source  
1..479  
/organism="Schistosoma mansoni"  
/mol\_type="genomic DNA"  
/strain="Puerto Rico"  
/db\_xref="taxon:6183"  
/clone="Sm1-50G22"  
/clone\_lib="Sm1"  
/note="Vector: pBelosBAC11; Site 1: Hin dIII; Constructed  
in the laboratory of Dr. Denis Le Paslier at the Fondation  
Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma  
mansoni agarose embedded DNA was partially digested with  
Hin dIII. High molecular weight fragments were ligated in  
pBelosBAC11 digested with Hin dIII. The average insert size  
is 100 kb. Total clone coverage: approx. 7.95 x the  
haploid genome. Further information can be found in Le  
Paslier et al. (2000) Construction and characterization of  
a Schistosoma mansoni bacterial artificial chromosome  
library. Genomics 65: 87-94."  
ORIGIN  
Query Match 4.0%; Score 60.8; DB 28; Length 479;  
Best Local Similarity 47.4%; Pred. No. 0.002;  
Matches 182; Conservative 0; Mismatches 202; Indels 0; Gaps 0;  
Qy 362 CAGCTAATACTACACAGTTTCCACTGTTACTGAGTCAAAATGATGCTAGGTTA 421  
Db |||||  
Qy 452 CTGCTGATGCTGCTGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 393  
Db |||||  
Qy 422 TTAAATGTTTCCAAATATGATATTAATGTTTCTAGCAGTGTAAATGAGCTGTAA 481  
Db |||||  
Qy 392 ATGATGATGCTGCCGCTGCTTATGATGATGATGATGATGATGATGATGATGATG 333  
Db |||||  
Qy 482 TTATGTTTACATCTGTAATCTCAATGCTACTATTTCATGAAAGAAATGATGATG 541  
Db |||||  
Qy 332 ATGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 273  
Db |||||  
Qy 542 GGCGAGATGCTGCTGGAAAACTGTAGATCAAAAAACGTTACTCGTTGGTGATACAGTCA 601  
Db |||||  
Qy 272 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 213  
Db |||||  
Qy 602 AATATACTATTACTTATAAGATGAGTCAATATATCATGGTACAGAAAGTGTATCAAT 661  
Db |||||  
Qy 212 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 153  
Db |||||  
Qy 662 ATGTTATAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 721  
Db |||||  
Qy 152 ATAAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 93  
Db |||||  
Qy 722 AAGTACTATTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 745  
Db |||||  
Qy 92 ACAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69  
Db |||||

```

Qy 493 TCTGTAAGTCCAAATCGTACTATTTCATGAAGAATACTCATGGAC 539
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 546 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 592

RESULT 21
BZ351620
LOCUS      BZ351620          525 bp     DNA         linear   GSS 12-NOV-2002
DEFINITION hw04f06.g1 WGS-SbicolorF (JM107 adapted methyl filtered) Sorghum
            bicolor genomic clone hw04f06 5', genomic survey sequence.
ACCESSION  BZ351620
VERSION    BZ351620.1 GI:24915123
KEYWORDS   GSS.
SOURCE     Sorghum bicolor (sorghum)
ORGANISM   Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE  1. (bases 1 to 525)
AUTHORS    Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B.R., Muller,S., Nascimento,L.,
            Zutavern,T., Palmer,L., McCombie,W.B. and Martienssen,R.A.
            Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
            Unpublished (2002)
COMMENT    Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mcombie@cshl.org
            Plate: hw04 row: f column: 06
            Seq primer: -2M13UnivRev
            Class: shotgun
FEATURES             Location/Qualifiers
     source           1..525
                     /organism="Sorghum bicolor"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4558"
                     /clone="hw04f06"
                     /lab_host="JM107 or DH5a"
                     /clone_lib="WGS-SbicolorF (JM107 adapted methyl filtered)"
                     /note="Site 1: Xba I; Site 2: Xba I; The vector was
                     digested with XbaI and one nucleotide was added by fill in
                     in the recessive 3' end. The genomic DNA was nebulized,
                     end repaired, adaptor ligated and size fractionated using
                     sephadex. The resulting fragments were between 0.8 and 3
                     kb and were cloned into the vector (.x/y reads in M13mp19,
                     .b/g reads in pUC19). The same ligation was transformed in
                     either JM107 or DH5a."

ORIGIN
Query Match      4.0%; Score 60; DB 28; Length 525;
Best Local Similarity 45.7%; Pred.No. 0.0031;
Matches 210; Conservative 0; Mismatches 250; Indels 0; Gaps 0;

Qy 80 CTTTTCGCCGTGAGCAGGGACAATTACAGTTCAAGATACTCAAAGGCGAACCTATA 139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9 CTACTGCTACTACTACTGCTGCTACTGCTGCTACTGCTGCTACTGCTGCTACTGCTACT 68
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 140 AAGCATATAAGPTTTTGATGCAGAAATAGATAAATGCAATGPATCTGATTCGAATAAG 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69 CTGCTACTACTGCTGCTACTGCTGCTACTGCTGCTACTGCTACTGCTACTGCTACTCTG 128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 200 ATGGAGGCTTCTTATTTAAATTCCTCAAGGTAAAAGAGCTGAGTATAAAGCTTCAACTGATT 259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 129 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 260 TTAAATCTCTTTTAGACAACTACTAATGGAGGGAGAACATATGTAACATAAAAAAGATA 319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

	RESULT	20	AZ183942	788 bp	DNA	linear	GSS 30-AUG-2000
	LOCUS		SP_1002_A1_H11_SP6 Strongylocentrotus purpuratus				
	DEFINITION		urChin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1002 Col=21 Row=O, genomic survey sequence.				
	ACCESSION		AZ183942				
	VERSION		AZ183942.1	GI:8356317			
	KEYWORDS		GSS.				
	SOURCE		Strongylocentrotus purpuratus				
	ORGANISM		Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euclinoidea; Echinacea; Echinoidea; Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 788)				
	REFERENCE		Cameron,R.A., Mahalras,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.  A sea urchin genome project: Sequence scan, virtual map, and additional resources Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000) 20402566 10920195				
	COMMENT		Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1002 row: O column: 21 Seq primer: SP6 Class: BAC ends High quality sequence stop: 788. Location/Qualifiers 1..788 /organism="Strongylocentrotus purpuratus" /mol_type="genomic DNA" /db_xref="taxon:7668" /clone="plate=1002 Col=21 Row=O" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library." /note="Organ: sperm; Vector: BCEJ3.6; BAC Clones in E-Coli DH10B"				
	FEATURES		source				
	ORIGIN		Query Match                4.0%; Score 60.6; DB 28; Length 788;				
			Best Local Similarity    50.4%; Pred. No. 0.0023;				
			Matches 175; Conservative    0; Mismatches 169; Indels    3; Gaps    1;				
	QY		193 AATAAGATGGAGCTTCTTAATTAAATTCCTCAAGGTAAAGAAGTGAGTTAAAGCTTCA	252			
	DB		249 AATAAATATGTTTTTTATGATGAATATGATGATCATGCGTGATGAAGATTATGAGGAG	308			
	QY		253 ACTGATTTTAA'T'CTCTTTTAGCACAACTACTAAATGAGGGAGAACATATGTACTAAA	312			
	DB		309 AA'TACTTGTC'T'CCTCCCTCTACTACTACTACTACTACTACTACTACTACTACTACT	368			
	QY		313 AAAAGATACTGC GT CAGCAAATGAGAT TGCACATGGGCTAAATCTATATCAGCTAACTACT	372			
	DB		369 ACT	425			
	QY		373 ACACCAGTTTCCA CTGTTACTGTAGTCAAAATAATGATGGTACTCAGGCTTAATATGTTTCC	432			
	DB		426 ATTACTACTACCA CCACCACCACCCTACTACTACTACTACTACTACTACTACTACTACT	485			
	QY		433 CAATA TGGA TAT TA TT ATGTTTCTAGCAC TGTTAAATGAGCTGTAATATGTTTACA	492			
	DB		486 ACT	545			

```
QY 320 CTGCGTCAGCAATGAGATTGGACATGGCTAAATCTATATACGCTAATACTACACAG 379
D 249 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 308
QY 380 TTTCCTACTGTTACTGAGTCAATTAATGATGGTACTGAGGTTAATTAATGTTTCCCAATATG 439
D 309 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 368
QY 440 GATATTATTATGTTCTAGACCTGTTTAATTAATGAGGCTGTAATTAATGTTTCAATCTGTAA 499
D 369 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 428
QY 500 CTCCAATGCTACTATTCTAGAAAGCAATGATGATGGAC 539
D 429 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 468

RESULT 22
CNS04AEE/c
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
095G24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL281759
ACCESSION
VERSION AL281759.1 GI:8020086
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.
1
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
PUBMED 10835645
REFERENCE 2
AUTHORS
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143
REFERENCE 3
AUTHORS
Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
Location/Qualifiers
1. 806
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="095G24"
/clone_lib="G"
/note="Genoscope sequence ID : CORG095BD12LP1-end : T7"

FEATURES
source
Query Match 4.0%; Score 60; DB 29; Length 806;
Best Local Similarity 41.6%; Pred. No. 0.0031;
Matches 218; Conservative 22; Mismatches 284; Indels 0; Gaps 0;

ORIGIN
```

```
QY 156 TGATGCAGAAATAGATAATGCAATGTATCTGATTCGAATAAAGATGGAGCTTCTTTATTT 215
D 753 TATTTACTACCAATATATATATATATATAGTAATATCATATAAATAATAGTACTACTAC 694
QY 216 AATTCCTCAAGGTTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAACTCTCTTTTAC 275
D 693 TACTACTAAATAATAATAATAGTAATAATAAATAATAATAATAATAATAATAATAATAATA 634
QY 276 GACAACTACTAATCGGAGGGAACATATGTAACCTAAAAAAGATACGTCGTCAGCAAAATGA 335
D 633 TAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAATAA 574
QY 336 GATTGCGACATGGGCTAAATCTATATCAGCTAATATACACACAGTTTCCACTGTCTACTGA 395
D 573 TAATAATGGTAATAGTAAMSSTATTATATVATACAAAHMACMMCMCMCMCMCMCMCM 514
QY 396 GTCAAAATAATGATGGTACTGAGGTTTATTAATGTTTCCCAATATGATATTTATGTTTC 455
D 513 YMAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATG 454
QY 456 TAGCACTGTTAATAATGAGGCTGTAATTTATGTTTACATCTGTAACCTCAAAATGCTACTAT 515
D 453 TAATGATAGTAATGATATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 394
QY 516 TCATGAAAAGAAATATCTGATCGACATGGGAGATGGTGGGAAAAAACTGTAGATCAAAA 575
D 393 TAATGGTAATAATAATGTTATTAATGATAGTAATSAVMATAATAGYMAVMAATAATAAG 334
QY 576 AACGTACTCGGTTGGTGTACAGTCAGTCAATATATCTTATTAAGAAATGCGAGTCAATTA 635
D 333 CAACCMCCACMATWATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAG 274
QY 636 TCATGCTGACAGAAAAGTGTATCAATATGTTTATAAGGATACTA 679
D 273 TAACGGTATTATGATAGTAATGATAATCATAGTAATAATAATAATA 230
```

## RESULT 23

CNS01VTG

LOCUS

DEFINITION

Tetraodon nigroviridis genome survey sequence T7 end of clone

199C24 of library G from Tetraodon nigroviridis, genomic survey

sequence.

AL169549

ACCESSION

VERSION AL169549.1 GI:7807606

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis

ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontoidea; Tetraodontidae; Tetraodon.

1

Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

Estimate of human gene number provided by genome-wide analysis

using Tetraodon nigroviridis DNA sequence

Nat. Genet. 25 (2), 235-238 (2000)

20296633

PUBMED 10835645

REFERENCE 2

AUTHORS

Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

Estimate of human gene number provided by genome-wide analysis

using Tetraodon nigroviridis DNA sequence

Nat. Genet. 25 (2), 235-238 (2000)

**TITLE** Direct Submission  
**JOURNAL** Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
**COMMENT** This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.

**FEATURES**

source

Location/Qualifiers  
1..773  
/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="199C24"  
/clone\_lib="G"  
/note="Genoscope sequence ID : COAG199BB12LPL-end : T7"

**ORIGIN**

Query Match 4.0%; Score 59.8; DB 29; Length 773;  
Best Local Similarity 41.7%; Pred. No. 0.0035;  
Matches 178; Conservative 26; Mismatches 223; Indels 0; Gaps 0;  
QY 200 ATGGAGCTCTTATTTAATCTCTCAAGGTAAAGAGCTGAGTATATAAGCTTCAACTGATT 259  
DB 136 ATCATGTTAATTTTATTTACTCATGATCTGCTACTACTACTACTACTACTACTACTG 195  
QY 260 TTAATCTCTTTTACGACAACTACTAATGGAGGAGAACATATGTAACTAAAAAGATA 319  
DB 196 CWACWACWCTACTGCTACTGCTACTACTAGTGTCTACTGCTACTGCTACTACTGCT 255  
QY 320 CTGCGTCAGCAATGAGATTGGCATGGGCTAAATCTATATCAGCTTAATCTACTACAC 379  
DB 256 CTACTACWAGTCTACTGCTACTGCTACTACTACTACTACTACTACTACTACTACTACT 315  
QY 380 TTTCACCTGTTACTGAGTCAATAATGATGGTACTGAGGTATTAACTGTTTCCCAATATG 439  
DB 316 CTRWACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 375  
QY 440 GATATTATTATGTTTCTAGCACTGTTAATAATGGAGCTGTTAATTATGTTTACATCTG 499  
DB 376 CATACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 435  
QY 500 CTCGAAATGCTACTATTATGAAGAATACTGATGGACATGGGAGATGGTGTGAA 559  
DB 436 CWACTACTACCACWACAACTACTCWCWRCAACTACTGCTACTACTACTACTACTACTC 495  
QY 560 AAATCTGTAGTCAAAAAAGTACTCGTTGGTGATACAGTCAATATATATCTATTCTTATA 619  
DB 496 CTGCTACTACTACTACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACT 555  
QY 620 AGAATGC 626  
DB 556 CTACTGC 562

**RESULT 24**

BJ391130

LOCUS

DEFINITION BJ391130 Dictyostelium discoideum cDNA library, SF Dictyostelium discoideum cDNA clone dds1506 5', mRNA sequence.

ACCESSION

BJ391130

VERSION

BJ391130.1

KEYWORDS

EST.

SOURCE

ORGANISM

Dictyostelium discoideum

Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.

AUTHORS

Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

TITLE

Full length cDNA of Dictyostelium discoideum at the slug stage

JOURNAL

Unpublished (2002)

COMMENT

Contact: Tadao Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

**FEATURES**

source

Location/Qualifiers  
1..617  
/organism="Dictyostelium discoideum"  
/mol\_type="mRNA"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="dds1506"  
/sex="mat A"  
/dev\_stage="slug stage"  
/clone\_lib="Dictyostelium discoideum cDNA library, SF"

**ORIGIN**

Query Match 3.9%; Score 59.6; DB 12; Length 617;  
Best Local Similarity 47.4%; Pred. No. 0.0038;  
Matches 179; Conservative 0; Mismatches 199; Indels 0; Gaps 0;  
QY 317 ATACTCGGTGAGCAAAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAATCTACAC 376  
DB 12 ATACTCAAACTACTACTACAACAACAACAATAATAATAATAATAATAATAATAATA 71  
QY 377 CAGTTTCCACTGTTCTAGTCAAAATGATGATGCTACTGAGGTTTAAATGTTTCCCAAT 436  
DB 72 ATATAATGATTTGATGATCAATGACTATAAATGATGCTGCTTCAACCTCTTCTACAG 131  
QY 437 ATGATATTTATTTGTTTCTAGCACTGTTAATGATGAGCTGTTAATGATGCTGTTACATCTG 496  
DB 132 TTGCTGAAGATTGCTAAGTACTACTATTGATGCTACAGCCACTATTACAGATCAACA 191  
QY 497 TAACTCCAAATGCTACTATTTCATGAAAAGAAATCTGATGCGACATGGGAGATGGTGGT 556  
DB 192 CTAATCATAAACATATGTTAGTGAAGCAATGCAAGATAGAGAATGAGAGATCAAAATA 251  
QY 557 GAAAACTGTAGTACAAAAACGTACTCGTGGTGTGATACAGTCAAAATATATCTATTCTT 616  
DB 252 ATAATGCAATGATGATAGTACAGATGAAATTAATAATATAGATAAAAAATAGTAATA 311  
QY 617 ATAAGAATGAGTCAATTTATCATGTCAGAAAAAGTGTATCAATATGTTTATAAAGGATA 676  
DB 312 ATATAGATAAAGAACAAGAGCAAGAGAAGAAGAAGAAGAAGAATAAATAAATAACA 371  
QY 677 CTATGCCATCTGCTTCTG 694  
DB 372 ATATAACATCAGATGATG 389

**RESULT 25**

BZ422321/c

LOCUS

DEFINITION BZ422321 WGS-SbicolorF (DH5a methyl filtered) Sorghum bicolor genomic clone id52c10 5', genomic survey sequence.

ACCESSION

BZ422321

VERSION

BZ422321.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Sorghum bicolor (sorghum)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE

1 (bases 1 to 427)

AUTHORS

Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N., Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L., Zucavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.

TITLE

Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)

JOURNAL

Unpublished (2002)

COMMENT

Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884



ORGANISM Strongylocentrotus purpuratus  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinoidea; Euechinoidea; Echinacea; Echinoida;  
Strongylocentrotidae; Strongylocentrotus.

REFERENCE 1 (bases 1 to 914)  
AUTHORS Cameron,R.A., Mahairas G., Rast,J.P., Martinez,P., Biondi,T.R.,  
Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T.,  
Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H.  
and Hood,L.  
TITLE A sea urchin genome project: Sequence scan, virtual map, and  
additional resources  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)  
MEDLINE 20402566  
PUBMED 10920195  
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91125, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: acameron@caltech.edu  
Plate: 100 row: M column: 24  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 914.

FEATURES source  
1..914  
Location/Qualifiers  
/organism="Strongylocentrotus purpuratus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7668"  
/clone\_lib="Plate=100 Col=24 Row=M"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea  
urchin, sperm genomic BAC library"  
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli  
DH10B"

ORIGIN  
Query Match 3.9%; Score 58.4; DB 28; Length 914;  
Best Local Similarity 50.0%; Pred. No. 0.0073;  
Matches 146; Conservative 0; Mismatches 146; Indels 0; Gaps 0;  
QY 248 CTTCACTGATTTTAACTCTCTTTTACGACACTACTAAATGGAGGAGACATATGTAA 307  
DB 167 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTATTA 226  
QY 308 CTAAGAAAGATCTGGTCAGCAATGAGATGGCAGATGGCTGAACTATATCAGCTA 367  
DB 227 CTACCACTACTACTACTGCGACTACTACTACTACTACTACTACTACTACTACTA 286  
QY 368 ATACTACCACTGTTCCACTGTTACTGAGTCAATTAATGATGGTACTGAGGTTATTAATG 427  
DB 287 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 346  
QY 428 TTCCCAATATGGATATTAATGTTCTAGCAGCTGTTAATTAATGAGCTGAATATGG 487  
DB 347 TTACTGTTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 406  
QY 498 TTACATCTGTAACCTCAATGCTACTATTTCATGAAAGAACTACTGTCGAC 539  
DB 407 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC 589

RESULT 28  
BH200391/c 548 bp DNA linear GSS 15-NOV-2002  
LOCUS Sml-39D16.TF Sml Schistosoma mansoni genomic clone Sml-39D16,  
DEFINITION genomic survey sequence.  
ACCESSION BH200391  
VERSION BH200391.1 GI:16370713  
KEYWORDS GSS.  
SOURCE Schistosoma mansoni  
ORGANISM Schistosoma mansoni  
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;

Strigoida; Schistosomatoida; Schistosomatidae; Schistosoma.  
1 (bases 1 to 548)  
Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and  
El-Sayed,N.M.  
TITLE Use of end sequences from Schistosoma mansoni (Puerto Rico strain)  
JOURNAL Sml BAC library for gene discovery and map construction  
COMMENT Unpublished (2001)  
Other GSSs: Sml-39D16.TR  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tigr.org  
Clones are derived from the Schistosoma mansoni (Puerto Rico) Sml  
BAC library. For clone availability, please contact Dr. Najib  
El-Sayed at TIGR (nelsayed@tigr.org) or Dr. Phillip LoVerde at  
State University of New York, Buffalo, New York, USA  
(loverde@buffalo.edu)  
Seq primer: M13 For  
Class: BAC ends.

FEATURES source  
1..548  
Location/Qualifiers  
/organism="Schistosoma mansoni"  
/mol\_type="genomic DNA"  
/strain="Puerto Rico"  
/db\_xref="taxon:6183"  
/clone="Sml-39D16"  
/clone\_lib="Sml"  
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed  
in the laboratory of Dr. Denis Le Paslier at the Fondation  
Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma  
mansoni agarose embedded DNA was partially digested with  
Hin dIII. High molecular weight fragments were ligated in  
pBelOBAC11 digested with Hin dIII. The average insert size  
is 100 kb. Total clone coverage: approx. 7.95 x the  
haploid genome. Further information can be found in Le  
Paslier et al. (2000) Construction and characterization of  
a Schistosoma mansoni bacterial artificial chromosome  
library. Genomics 65: 87-94."

ORIGIN  
Query Match 3.8%; Score 58; DB 28; Length 548;  
Best Local Similarity 46.9%; Pred. No. 0.0089;  
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0;  
QY 339 TGCACATGGGCTAAATCTATATCAGCTAATACACACAGTTCCACTGTTACTGATC 398  
DB 514 TGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 455  
QY 399 AAATAATGATGGTACTGAGGTTAATATGTTTCCCAATATGGATATTATATGTTTCTAG 458  
DB 454 TGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 395  
QY 459 CACTGTTAATTAATGAGCTGTAATTAATGTTTACATCTGTAACTCAAAATGCTACTATTCA 518  
DB 394 TGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 335  
QY 519 TGAAGAAGAACTGATGCGACATGGGAGATGGTGGTGGAAAACTGTAGATCAAAAAC 578  
DB 334 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 275  
QY 579 GTACTCGTGGTATACAGTCAATATATCTATTACTTATAAGAAATGCAGTCAATTATCA 638  
DB 274 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 215  
QY 639 TGGTACAGAAAAAGTGTATCAATATGTTTAAAGGATACCTATGCCATCTGCTTCTGTAGT 698  
DB 214 TGAAGAAGATGATGCTGGATGAAGTTGAAGAAAGGATGATAAAGATTTTGGAAATGATGT 155  
QY 699 TGATTTGAACGAAGGGTCTTATGAAG 724

Db	154	GTTTTGTGACGATGACGATGATGTAG	129
RESULT 29			
FR0006944			
LOCUS			
DEFINITION	619 bp	DNA	linear
ACCESSION	FR0006944		GSS 02-MAR-1997
VERSION	290754.1		GI:1867968
KEYWORDS	GSS; genome survey sequence.		
SOURCE	Takifugu rubripes (Fugu rubripes)		
ORGANISM	Takifugu rubripes		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.		
1	(bases 1 to 619)		
Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrانيا,Y., Williams,G. and Brenner,S.			
Direct Submission			
Submitted (18-FEB-1997)			MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmrc.ac.uk
Vector: pBluescript II KS			
V type: phagemid			
PRIMER: KS			
DESCR:			
One pass dye-terminator sequencing of cosmid cloned genomic sequence.			
FEATURES			
source			
1..619			
/organism="Takifugu rubripes"			
/mol_type="genomic DNA"			
/db_xref="taxon:31033"			
/clone_lib="133B16aC7"			
/clone_lib="cosmid 133B16"			
ORIGIN			
Query Match	3.8%;	Score 58;	DB 29; Length 619;
Best Local Similarity	46.7%;	Pred. No. 0.0089;	
Matches 169;	Conservative	0; Mismatches 193;	Indels 0; Gaps 0;
Qy	156	TGATGCGAAGATAGATAATGCAATGATCTGATTCGAATAAGATGGAGCTTCATTATTT	215
Db	14	TGCTACTANNACTGCTGCTACNGCTACTACNGCTNCTGCTACTAATGCTACTGCTGCTAC	73
Qy	216	AATTCCTCAAGGTAAGAGCTGAGTAAAGCTTCAAGCTGATTTAAATTCCTCTTTTAC	275
Db	74	TGCNACTACTGCTACTGCTGCTACTACTGCTACTGCTGCTACTACTACTACTGCTGCTG	133
Qy	276	GACAACTACTAATGGGAGGAGACATATGTAATCAATAAAAAAGATACGCGTCAGCAATGA	335
Db	134	TGCTGCTGCTGCTGCTACTACTGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	193
Qy	336	GATTGCGACATGGGCTAAATCTATATCAGCTAATATACACAGTTTCCACTGTTACTGA	395
Db	194	TACTACTACTGCTGCTACTACTGCTGCTGCTGCTGCTACTACTGCTGCTGCTACTACT	253
Qy	396	GTCAATAATGATGCTACTGAGGTTTAAATGTTTCCCAATATGGAATATTATTATGTTTC	455
Db	254	TACTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTACTACTGCTGCTGCTG	313
Qy	456	TAGCACTGTTTAAATGAGCTGTAATATGTTTACATCTGTAACTCCAAATGCTACTAT	515
Db	314	TACTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	373
Qy	516	TC 517	
Db	374	TC 375	
RESULT 30			
CE682324			
LOCUS			
DEFINITION	641 bp	DNA	linear
ACCESSION	CE682324		GSS 29-SEP-2003
VERSION	290754.1		GI:1867968
KEYWORDS	GSS; genome survey sequence.		
SOURCE	Takifugu rubripes (Fugu rubripes)		
ORGANISM	Takifugu rubripes		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.		
1	(bases 1 to 641)		
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.			
The dog genome: survey sequencing and comparative analysis			
Science 301 (5641), 1898-1903 (2003)			
22875432			
14512627			
Contact: Kirkness EF			
The Institute for Genomic Research			
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,			
Rockville, MD 20850, USA			
Tel: 301-838-0200			
Fax: 301-838-0208			
Email: ekirknes@tigr.org			
Class: shotgun.			
FEATURES			
source			
1..641			
/organism="Canis familiaris"			
/mol_type="genomic DNA"			
/strain="Standard Poodle"			
/db_xref="taxon:9615"			
/clone_lib="Dog Library"			
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"			
ORIGIN			
Query Match	3.8%;	Score 57.6;	DB 29; Length 641;
Best Local Similarity	48.7%;	Pred. No. 0.011;	
Matches 187;	Conservative	0; Mismatches 194;	Indels 3; Gaps 1;
Qy	153	TTTTCATGACGAAATAGATAATGCAATGATCTGATTCGAATAAGATGGAGCTTCCTTA	212
Db	261	TGTTACTGAGATGATGATGCTACAAATACTACTGCTGCTGTTATTATTAATGCTATGTA	320
Qy	213	TTTAATTCCTCAAGTAAAGAGCTGAGTATAAGCTTCAAGCTTCAATTTTAAATTCCTTTT	272
Db	321	TGCTACAGCTACTGCTAATGCTGCTGCAATAACACTAATACCACTGCTGCTGCTATCAT	380
Qy	273	TACGCAACTACTAATGAGGAGGAGACATATGTAATAAAAAAGATACGCTGCTGAGCAA	332
Db	381	TACTAGTGTAGTGTGATGCTACTACCATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCT	440
Qy	333	TGAGATTGCGACATGGGCTAAATCTATATCAGCTAATATACACAGTTTCCACTGTTTAC	392
Db	441	TAAATGCTACTAATGCTACTGCTACCACTCATATGCTAGTGTGCAACTCTCTGAAAT	500
Qy	393	TGAGTCAATAATGATGCTGAGGTTTAAATGTTTCCCAATATGATATATATGCT	452
Db	501	ATCTACTAATACTAGTGTGCTAATGCTGCTAACAATACTGTTGGAATCTAATGCTGTA	560
Qy	453	TTCTAGCACTGTTAATGAGGCTGTTAATGTTTACATCTGTAACCTCCAAATGCTTAC	512
Db	561	TACTGCTGATATCAATGCACTACTGATGTTACAGCT---AATGTTAATGCTAC	617
Qy	513	TATTCATGAAAGAACTACTGATGC	536
Db	618	TGGAATACAACTAATGCTGCTGC	641

Search completed: July 27, 2004, 01:14:14  
Job time : 4191 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2004, 22:21:53 ; Search time 124 Seconds  
(without alignments)  
6753.394 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaagaatgattca.....gtcgtcggttacgttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64.8	4.3	1980	4	US-09-134-000C-2719
2	64.8	4.3	1983	4	US-09-071-035-217
3	56.2	3.7	1575	4	US-09-543-681A-615
4	50.4	3.3	14066	4	US-09-601-198-56
5	49.8	3.3	1687	4	US-09-071-035-219
6	48.4	3.2	3666	2	US-08-682-517-13
7	48.4	3.2	3666	2	US-08-682-517-14
8	48.4	3.2	4197	2	US-08-682-517-7
9	48.4	3.2	4197	2	US-08-682-517-8
10	47.8	3.2	740	3	US-09-451-117-1
11	47.8	3.2	740	4	US-09-888-655-1
12	47.4	3.1	1037	4	US-09-181-585-3
13	47.4	3.1	1159	4	US-09-181-585-1
14	47.4	3.1	1471	4	US-09-181-585-2
15	47.4	3.1	3945	4	US-09-200-6508-6
16	47.4	3.1	4358	4	US-08-956-171E-454
17	47.2	3.1	7218	1	US-08-232-463-14
18	46.8	3.1	1948	1	US-07-849-438-1
19	46.8	3.1	4590	4	US-09-134-001C-1108
20	45.6	3.0	4376	1	US-08-119-125A-1
21	45.6	3.0	6744	1	US-08-119-125A-2
22	45	3.0	832	4	US-09-621-976-2813
23	44.6	3.0	11679	4	US-09-328-352-1377
24	44.2	2.9	606	4	US-09-601-198-166
25	44.2	2.9	30549	4	US-09-134-001C-322
26	44	2.9	1716	4	US-09-134-001C-1028
27	43.4	2.9	2763	4	US-09-463-402-5

28	43.4	2.9	2766	4	US-09-889-572-3	Sequence 3, Appli
29	43.4	2.9	4065	4	US-09-463-402-13	Sequence 13, Appl
30	43.2	2.9	5181	1	US-08-257-073-10	Sequence 10, Appl
31	43	2.8	2001	4	US-09-528-784A-84	Sequence 84, Appl
32	43	2.8	2001	4	US-09-569-058A-84	Sequence 84, Appl
33	43	2.8	3057	4	US-09-601-198-55	Sequence 55, Appl
34	43	2.8	3402	4	US-09-528-784A-86	Sequence 86, Appl
35	43	2.8	3402	4	US-09-569-058A-86	Sequence 86, Appl
36	42.8	2.8	966	4	US-09-134-001C-1789	Sequence 1789, Ap
37	42.6	2.8	168575	4	US-09-426-290-1	Sequence 1, Appli
38	42.2	2.8	810	4	US-09-601-198-31	Sequence 31, Appl
39	42.2	2.8	3292	1	US-07-814-964-12	Sequence 12, Appl
40	42.2	2.8	3292	1	US-08-258-442-12	Sequence 12, Appl
41	42.2	2.8	3292	1	US-08-328-809-7	Sequence 7, Appli
42	42.2	2.8	3292	4	US-08-866-840-7	Sequence 7, Appli
43	42.2	2.8	3292	5	PCT-US92-11107-12	Sequence 12, Appl
44	42.2	2.8	4185	4	US-09-417-485D-7	Sequence 7, Appli
45	42.2	2.8	10640	4	US-09-417-485D-5	Sequence 5, Appli
46	42	2.8	6060	4	US-08-956-171E-534	Sequence 534, App
47	41.8	2.8	1581	4	US-09-601-198-80	Sequence 80, Appl
48	41.6	2.8	1088	4	US-09-175-684A-2	Sequence 2, Appli
49	41.6	2.8	1842	4	US-09-117-415B-21	Sequence 21, Appl
50	41.6	2.8	1896	4	US-09-117-415B-17	Sequence 17, Appl
51	41.6	2.8	1896	4	US-09-117-415B-19	Sequence 19, Appl
52	41.6	2.8	1950	4	US-09-117-415B-15	Sequence 15, Appl
53	41.4	2.7	429	4	US-09-543-681A-707	Sequence 707, App
54	41.4	2.7	1555	3	US-08-669-408B-9	Sequence 9, Appli
55	41.4	2.7	1881	4	US-09-328-352-3710	Sequence 3710, Ap
56	41.4	2.7	5340	4	US-09-627-122-21	Sequence 21, Appl
57	41.4	2.7	580073	4	US-08-545-528D-1	Sequence 1, Appli
58	41	2.7	4860	4	US-09-328-352-3221	Sequence 3221, Ap
59	40.8	2.7	1431	3	US-09-316-083-2	Sequence 2, Appli
60	40.8	2.7	1431	4	US-09-933-700-2	Sequence 2, Appli
61	40.8	2.7	1806	4	US-09-134-001C-954	Sequence 954, App
62	40.8	2.7	4865	3	US-08-894-017-24	Sequence 24, Appl
63	40.8	2.7	4865	4	US-09-456-474-24	Sequence 24, Appl
64	40.8	2.7	5177	6	5352450-1	Patent No. 5352450
65	40.8	2.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
66	40.6	2.7	414	2	US-08-630-822A-63	Sequence 63, Appl
67	40.6	2.7	414	2	US-09-005-069-63	Sequence 63, Appl
68	40.6	2.7	414	4	US-09-171-156A-22	Sequence 22, Appl
69	40.6	2.7	414	4	US-09-004-730A-22	Sequence 22, Appl
70	40.6	2.7	414	4	US-08-981-799A-22	Sequence 22, Appl
71	40.6	2.7	567	4	US-09-543-681A-409	Sequence 409, App
72	40.6	2.7	3255	4	US-09-351-198-1	Sequence 108, App
73	40.6	2.7	4291	4	US-09-351-200-1	Sequence 1, Appli
74	40.4	2.7	363	4	US-09-328-352-2509	Sequence 2509, Ap
75	40.2	2.7	2202	4	US-09-328-352-1473	Sequence 1473, Ap
76	40.2	2.7	3892	3	US-08-569-214-1	Sequence 1, Appli
77	40.2	2.7	3892	3	US-08-937-236-1	Sequence 1, Appli
78	40	2.7	291	4	US-09-134-001C-1848	Sequence 1848, Ap
79	40	2.7	1197	4	US-09-134-001C-1887	Sequence 1887, Ap
80	40	2.7	1785	4	US-09-117-415B-1	Sequence 1, Appli
81	40	2.7	4605	4	US-09-543-681A-1010	Sequence 1010, Ap
82	39.6	2.6	1338	1	US-08-485-216-1	Sequence 1, Appli
83	39.6	2.6	1338	2	US-09-003-245-1	Sequence 1, Appli
84	39.6	2.6	1338	3	US-08-853-552-1	Sequence 1, Appli
85	39.6	2.6	2112	4	US-08-956-171E-487	Sequence 487, App
86	39.4	2.6	417	4	US-09-107-532A-3029	Sequence 3029, Ap
87	39.4	2.6	1101	4	US-09-134-001C-1575	Sequence 1575, Ap
88	39.4	2.6	1104	4	US-09-134-001C-841	Sequence 841, App
89	39.4	2.6	1134	4	US-09-540-236-1714	Sequence 1714, Ap
90	39.4	2.6	2237	4	US-08-914-999-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-09-134-000C-2719  
; Sequence 2719, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:

```

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2719
; LENGTH: 1980
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; US-09-134-000C-2719

Query Match          4.3%; Score 64.8; DB 4; Length 1980;
Best Local Similarity 52.1%; Pred. No. 9.4e-07;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

Qy 1180 ACAGAAGCTAATGCAACAGAAATATACACAGGAGCAGATGGTATTAATTTACCATTTACAGGC 1239
      |||||
Db 1657 ACAAAGCTGAAGCAACTACTTTTACAAACCGCTGATGGATTAGTTGATATACACAGG 1716

Qy 1240 TTCAAGAGAGGTACATACTACTCTAGTTGAGAAAAGGCTCCCTTAGTTTACAATTTGTGTTA 1299
      |||||
Db 1717 CTTAAATACGGTACCTATTATTATTAGAAAGAACTGTAGCTCTCGATGATTATGTCTTGTTA 1776

Qy 1300 GATAACTCTCAGAAGGTTATTTTATAGGAGATGGAGCCACTGTATACGACTAAATTCAGATAAC 1359
      |||||
Db 1777 ACAATCGGATTGAAATTTGTGGTCAATGAACAATCATATGGCACAA-----CAGAAAAC 1830

Qy 1360 CTTTTAGTTAA CCGAACTGTTGAAAATAACAAAGGTACTGTAGTTGCTTTCACAGGTGGT 1419
      |||||
Db 1831 CTAGTTTTCACGAAAAGAGTACCAACAAACAAAGGTACCTTACCTTCAACAGGTGGC 1890

Qy 1420 ATTGGTACACAAATTTCTACATTATAGTGTCAAATTTTAGTAAATAGGAGCAGGTATCGTG 1479
      |||||
Db 1891 AAGGAATCTACGTTTACTTGAAGTGGCGAGCTTGCTACTATTATGCAGGAGTCTAC 1950

Qy 1480 CTTGTTGCTCGTCGTCTGTTTACGTTCTTAA 1509
      |||||
Db 1951 TTTGCTAGACCTTAGAAAAGAAAATGCTTAA 1980

```

RESULT 2  
US-09-071-035-217  
; Sequence 217, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:

```

; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 217:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1983 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-09-071-035-217

Query Match 4.3%; Score 64.8; DB 4; Length 1983;
Best Local Similarity 52.1%; Pred. No. 9.4e-07;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAAGCTAATGCACAGAAATATACACAGGAGCAGATGGTATTAATTACCATTCACAGC 1239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1660 ACAAAAAGCTGAACCACTACTTTTACAACAACGGCTGATGGATTAGTTGATATCACAGGG 1719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1240 TTCAAAAGAGGTACATACTATCTAGTTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTGA 1299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1720 CTTTAATACGGTACCTATTATTTTAGAAGAAACTGTAGTCTCTGATGATTATGCTTGTGA 1779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1300 GATAACTCTCAGAAGGTTATTTTAGAGATGGAGGCACCTGATACGACTAATTCAGATAAC 1359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1780 ACAAAATCGGATTGAAATTTGTGGTCAATGAACAATCATATGGCACAA-----CAGAAAAC 1833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1360 CTTTTAGTTTAACCCAACTGTTGAABAATAACAAGGTACTGAGTTGCCTTCAACAGGTGGT 1419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1834 CTAGTTTTCACCGAAGAAAGTACCAAAACAAACAAAGGTACCTTACCCTTCAACAGGTGGC 1893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1420 ATTGGTACAACAATTTCTACATATATAGGTGCAATTTTAGTAATAGGACGAGTATCGTG 1479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1894 AAGAGAACTACGCTTTACTTAGGAAGTGGCGCAGTCTTGCTACTTATTGCGAGGACTAC 1953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1480 CTTGTTGCTCGTCGTCTGTTTACGTTCTTAA 1509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1954 TTTGCTAGACGTAGAAAAGAAAATGCTTAA 1983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 3
US-09-543-681A-615
; Sequence 615, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEIN
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 615
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-615

```

	Query Match	3.7%	Score 56.2;	DB 4;	Length 1575;
	Best Local Similarity	43.3%;	Pred. No. 0.00012;		
	Matches 262;	Conservative	0;	Mismatches 343;	Indels 0; Gaps 0;
Qy	107	CAGTTCAGGATCTCAAAAAGGGCAACCTATAAAGCATATAAAGTTTTGATGCAGAA	166		
Db	404	CAATTTTTCTAATGAANAATTCGATTGNAATGCTCTATATGAAGAAGTTAGATAAAAATGAGC	463		
Qy	167	TAGATAATGCAAAATGATCTGATTCGAATAAAGATGGAGGCTCTTTATTTTAATTCCTCAAG	226		

464	TTAATGATCAAAATTTTGGCTGAGTGTGATATATGAAATTAAGATAATAAATACGTTCTGGT	523
227	GTAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTCCTTTTTACGCAACTACTA	286
524	AAATAGACTTTAAAGCTTAATAAAGTAACCTTTAATCAAGTTAACTTTTAGTCAATAAGAC	583
287	ATGAGGGAGAACATATGTAACTTAAATAAAGATACCTGGCTCAGCAAAATGAGATTCGGACAT	346
584	AGCTCGTAAAGCAAGCGATAGAAAATGATAATACTTAACCTTAGATACAGAAAAATGATAATA	643
347	GGGCTAAATCTATATCAGCTAAATACTACACCAGTTTCCACCTGTTACTCGAGTCAAAATATG	406
644	TTAACTTAGAAAATGATAATTAACCTTAGAAAATGATAATTAATTAACCTTAGAAAATGATA	703
407	ATGCTACTGAGGTTTAAATGTTTCCCAATAATGGATATTATATGTTTCTAGCACCTGTTA	466
704	ATATTAACTTAGAAAATGATAATTAATTAACCTTAGAAAATGATAATTAATTAACCTTAGAAAATG	763
467	ATAATGAGCTGTAATTTATGGTTACATCTCTGTAACCTCCAAATGCTACTATTTCATGAAAAAGA	526
764	ATAATATTAACTTAGAAAAATGATAATATTAACTTTAGAAAAATGATAATGTTTAACTTAGAAA	823
527	ATACTGATGGACATGGGGAGATGGTGGTGGAAAAAATCTGTAGATCAAAAAACGTACTCGG	586
824	ATGATAATGTTAACTTTAGAAAATGATAATGTTTAACTTTAGAAAATGATAATGTTTAACTTTAG	883
587	TTGGTGATACAGTCAAAATATACTATTACTTTATAAGAATGCAAGTCAATTAATCATCGTACAG	646
884	AAAATGATAATGTTTAACTTAGAAAATGATAATATTAACTTAGAAAATGATAATGCTTAAC	943
647	AAAAAGTGATCAATATGTTTATAAGGATACTATGCCATCTGCTTCTGCTAGTTGATTTGA	706
944	TAAATATAAAAAGGCCTAATAAAGATATAAATAGAAATACTTATATTTTCAATGGTA	1003
707	ACGAA	711
1004	AAGAA	1008

RESULT 4  
US-09-601-198-56/c

03-03-001-138-307C  
; Sequence 56, Applica  
; Patent No. 6531583  
; GENERAL INFORMATION:

	Query Match	3.3%	Score 50.4	DB 4	Length 14066
	Best Local Similarity	42.8%	Pred. No. 0.007		
	Matches 313	Conservative 0	Mismatches 416	Indels 3	Gaps 1
Qy	435	ATATGGGATATTATTATGTTTCTAGCACCTGTTTAAATAATGGAGCTGTAATTATGTTTACATC	494		
Db	5284	AAATGTTGAGTTCTAATGATCTTATAAATACATCAACACAAACAGTTAATTATGATTATGA	5225		

Qy	495	TGTAATCCAAATGCTACTATTATCAAGAAAAGAAATATCTGATCGGCATATGGGGAGATGTGGTGG	554
Db	5224	TGGAAATTCACGCAACTTGAATTAATAAAAAAATTCAAATTTGGTATATATCTTCCAATGACGG	5165
Qy	555	TGCAAAATCTGTAGATCAAAAACGTPACTTCGGTTCGTGATACAGTCAAAATATATCTATTAC	614
Db	5164	TGAAGAAATTTTAAGTGATCAAAAACAATTCGCTTTGGGCTAATAACCACTATAGTTTGA	5105
Qy	615	TTATAAGAATGCAAGTCAATTTATCATCGGTACAGAAAAAGTGTATCAATATGTTTATAAAGGA	674
Db	5104	ATTGTCCTAATTTAAACATAAATCGTAAATATACATTAAGAAAGTAAGAAATTAATTAATGA	5045
Qy	675	TACTATGCCATTCGCTTCCTGTAGTTTGAACGAAGGCTCTTATGAAGTAACCTATTATAC	734
Db	5044	TGATAATAAAACATCAATTAATTTTCCATTTAAAAAAATGGAATAGCTGATTTGATTTATTTGT	4985
Qy	735	TGATGGATCAGGGAATATTACAACTCTAACTCAAGGTTTCGGAAGAAAGCAACTGGGAAGTA	794
Db	4984	TAATAAAACACAATTCGATTAGTATTAAGTTCAATTAATTTGAGCCCAATTTGCAAGAGCGGAAAA	4925
Qy	795	TAACTCTTTAGAGGAAAAATAAATTTTCAAGATTACTATTCCTGGTGGGAGCTACCAATAC	854
Db	4924	TAATTTACAATCAACACAATTTAGATTTATATTAATGATCCTGATAATGTTTAAAGTAA	4865
Qy	855	TCCAACCGGAATPACTCAAAATGGAGCTAATGATGACTTTTTTTTATAAGGGAATAAATAC	914
Db	4864	TGAAGAAGAAGCAATCATTTAATTTATGGTGAGAAT...TTATCTGTAAAGCAAAAAGTTAA	4808
Qy	915	AATCACAGTCACTTATACAGGAGTATTTAAAGAGTGGAGCTAAACCAAGGTTTCAGCTGATTT	974
Db	4807	AATTTGTAGGAGTCAAAAATATTTTAGAAAACAATTTCAATTAATCTAGCAATTTAAACCAAGA	4748
Qy	975	ACCAGAAAAATACAAACATTTGGCAGCACTCAACCCCAATCTACCAATCTGATGCCAGGTCA	1034
Db	4747	TACTATTATTAACAGTATTACTTTTATTAATAAACAATAAAGCTGCTACGAATATTGG	4688
Qy	1035	AAAAGTAACAGTGAGGGATGGTCAAATTACTATAAAAAAAATTTGATGTTCCACAAAAGC	1094
Db	4687	AATAGATAATAGCAATAAATAATATAAACAATACAAACAATAATAAATCCATTTAAAAATTGA	4628
Qy	1095	TTCAATTAACAGGTGCTATATTGTTTTTAAAGAAATGCTACGGGTCAATTTCTAAACTTTAA	1154
Db	4627	TAATAATTTTAGTGCTCAITTTGGTCTGCTTCCAAAAGTGATACTAATAAACGTTATTTTGA	4568
Qy	1155	CGATACAAATAA	1166
Db	4567	ATATAATAAA	4556

RESULT 5

US-09-071-035-219

; Sequence 219, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/071,035

; FILING DATE:

; CLASSIFICATION:





QY	662	ATGTTATAAAGGATACTATGCCATCTGCTTCTGTAGTTGATTTTCAACGAAGGGTCTTATG	721
Db	3312	TAAATATTGATTAATGCTGGTGCTCAAGTAATTAACTTAGCAGGTAAAAAGGTGCACAAG	3371
QY	722	AAGTAACTATTACT	735
Db	3372	GTGTAGCTGATGCT	3385

```

RESULT 10
US-09-451-117-1
; Sequence 1, Application US/09451117
; Patent No. 6277973
; GENERAL INFORMATION:
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Fayer, Ronald
; APPLICANT: Trout, James
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein
; FILE REFERENCE: 0046.99
; CURRENT APPLICATION NUMBER: US/09/451,117
; CURRENT FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-09-451-117-1

```

Query Match	3.2%	Score 47.8	DB 3	Length 740	
Best Local Similarity	46.7%	Pred. No. 0.011			
Matches 151	Conservative	0	Mismatches 172	Indels 0	Gaps 0
QY	217	ATTCCTCAAGGTAAGAAGCTGAGTATATAAGGCTTCAACTGATTATTAATCTCTCTTTTACG	276		
Db	370	ACTACTCNAATCAATGGAAAGTTACTGATACTAATATTGGTAATATGAGTGGTATTATTACA	429		
QY	277	ACAACACTAATATGGAGGGAGAACATATGTATCTAAAAAGATATCTGCTCAGCAAAATGAG	336		
Db	430	TCTAGTGGTGATTCCTATAGCTGTTACTAATAATCTTAATGGTAATAATAATAGTAATAGT	489		
QY	337	ATTGGGACATGGGCTAAATCTATATCAGCTAATACTACACAGGTTTCCACTGTTTACTGAG	396		
Db	490	AATATTGGATCAGGAAATTTTATACCAGTTGGTACTGTTCTTCTTACTAGTATTGGTAAAT	549		
QY	397	TCAATAATATGATGGTACTGAGGTATTAATGTTTCCCAATATGGATATATTATGTTTCT	456		
Db	550	AGTAATGGTGTGCTTTTACTGCTAATTCATCCTTAATAATAACAATGACAAATAATTAAAT	609		
QY	457	AGCACTGTTTAATAATGGAGCTGTAAATTATGGTTACATCTGTAACCTCCAAATGCTACTAAT	516		
Db	610	AATATAATAATAATAATAGTAATAACCACCTCTTACTACTGTTGCTACTAATGCTAAATAT	669		
QY	517	CATGAAAAAGAACTGATGGCGAC	539		
Db	670	ACTACTAATACTACTAAATACTAC	692		

```

RESULT 11
US-09-888-655-1
; Sequence 1, Application US/09888655
; Patent No. 6521229
; GENERAL INFORMATION:
; APPLICANT: Jenkins, Mark C.
; APPLICANT: Payer, Ronald
; APPLICANT: Trout, James
; TITLE OF INVENTION: Cloning and Expression of a DNA Sequence Encoding A 41
; Patent No. 6521229
; TITLE OF INVENTION: kDa Cryptosporidium parvum Oocyst Wall Protein
; FILE REFERENCE: 0046.99

```

```

; CURRENT APPLICATION NUMBER: US/09/888,655
; CURRENT FILING DATE: 2001-06-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1:
; LENGTH: 740
; TYPE: DNA
; ORGANISM: Cryptosporidium parvum
US-09-888-655-1

Query Match      3.2%; Score 47.8; DB 4; Length 740;
Best Local Similarity 46.7%; Pred. No. 0.011;
Matches 151; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 217 ATTCTCAAGGTAAAGAAGCTGAGTATAAAGCTTCAACGATTTTAACTCTCTTTTACG 276
DB 370 ACTACTCAATCAATGGAAGTTACTGATCTACTAATATTGGTAATATGAGTGGTATTATTACA 429
QY 277 ACAACTACTAATGCGGGAGAACATATGTAACATAAAAAACATACTCGTCAGCAAAATGAG 336
DB 430 TCTAGTGGGTACTTATAGCTGTGTACTTAATACTTAAUGGTAATAATAATAGTAATAGT 489
QY 337 ATTGCGACATGGGCTAAATCTATATCAGCTAAATACTACACAGTTTCCACTGTTTACTGAG 396
DB 490 AATATTGGATCAGGAATTTTATACCAGTTGGTACTTGTCTTCTACTAGTATTGGTAAT 549
QY 397 TCAATAAATGATGGTACTGAGGTTATTAATGTTTCCCAATATGGAATATTATATGTTTCT 456
DB 550 AGTAATGGTGTGCTTTTACTGCTATTTCATCCTAATAATAACAATAGCAATAATATTAAAT 609
QY 457 AGCACTGTTAATAATGAGGCTGTAAATATGTTGATACATCTGTAACCTCAAAATGCTACTATT 516
DB 610 AATAAATAATAATAATAATAGTAATATACCACTCTTACTACTGTTGCTAATATGCTAATATT 669
QY 517 CATGAAAGAATACTGATGCGAC 539
DB 670 ACTACTAATACTACTAATACTAC 692

RESULT 12
US-09-181-585-3
; Sequence 3, Application US/09181585
; Patent No. 6524791
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 1100900101
; CURRENT APPLICATION NUMBER: US/09/181,585
; CURRENT FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1037
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: comprising exons E, C, and A
; US-09-181-585-3

```

[illegible]







[illegible]

RESULT 18  
 US-07-849-438-1  
 ; Sequence 1, Application US/07849438  
 ; Patent No. 5459034  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TABAQCHALI  
 ; TITLE OF INVENTION: C. Difficile Specific Oligonucleotides  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: 3i Research Exploitation Limited  
 ; STREET: The Gate House, 2 Park Street  
 ; CITY: Windsor  
 ; STATE: Berkshire  
 ; COUNTRY: U.K.  
 ; ZIP: SL4 1LU  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/849,438  
 ; FILING DATE: 19920519  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: WO PCT/GB90/02031  
 ; FILING DATE: 28-DEC-1990  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 8929293.2  
 ; FILING DATE: 29-DEC-1989  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1948 base pairs

```

; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Clostridium difficile
; US-07-849-438-1

Query Match          3.1%; Score 46.8; DB 1; Length 1948;
Best Local Similarity 45.4%; Pred. No. 0.027;
Matches 168; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

Qy 185 CTGATTCGAATAAAGATGGAGCTCTTTATTTATTCCTCAAGGTAAGAAGCTGAGTATA 244
    |||||
Db 794 CTGATCATAATAATAGAAAGGTCAGGCTATAGTTTATCAAGTAAATCTTAACTTTGA 853
    |||||

Qy 245 AAGCTTCAACTGATTTTTAAATCTCTTTTACGACAACTACTAATGAGGGGAGAACATATG 304
    |||||
Db 854 ATGGTAAAAAATATATTTTGTAAATAAATCAAAAGCAGGTTACTGGATGGCAAACTATTG 913
    |||||

Qy 305 TAACCTAAAAAGATACCTGCTGACGAAATGAGATTCGACATGGGCTTAAATCTATATCAG 364
    |||||
Db 914 ATAGTAAAAAATATTACTTTTAATTACTAACTGCTGAAGCAGCTACTGGATGGCAAACTA 973
    |||||

Qy 365 CTAATACTACACCGATTTCCACTGTTTACTGAGTCAAAATAATATGATGGTACTGAGGTTATTA 424
    |||||
Db 974 TTGATGGTAAAAAATATTACTTTTAATACTAACTGCTGAAAGCAGCTACTCTGGATGGCAA 1033
    |||||

Qy 425 ATGTTTTCCCAATATGGATATATTATGTTTCTAGCACTGTTTAATAATGGAGCTGTAATTA 484
    |||||
Db 1034 CTATTGATGTGAAAAAATATTACTTTTAATACTAACTGCTATAGCTTTCAACTGGTTTATA 1093
    |||||

Qy 485 TGGTTTACATCTGTAACCTCCAAATGCTACTATTATGAAAAAATACTGATGCCACATGGG 544
    |||||
Db 1094 CAATTATTAATGTTAAACATTTTTTATTTTAATCTAGTATGTTATATCGGAGTGTTTAAG 1153
    |||||

Qy 545 GAGATGGTGG 554
    |||||
Db 1154 GACCTAATGG 1163
    |||||

```

```

RESULT 19
US-09-134-001C-1108
; Sequence 1108, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHY
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1108
; LENGTH: 4590
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1108

Query Match          3.1%; Score 46.8; DB 4; Length 4590;
Best Local Similarity 47.6%; Pred. No. 0.037;
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

Qy      289  GGAGGGAGAACATATGTAACTAAAAAGATACTCGGTCAGCAAAATGAGATTGGCAGCATGG 348
Db      2899  GAAGGTGGCAACACTACCAGGTTCAAAAGTATATCCATTGGATTGCTGAGACTAGAGGTTGG 2958
Qy      349  GCTAAATCTTATATCAGCTAATACTACACCAAGTTTCCCACTGTTTACTGAGTCAAAATATGAT 408

```

Db 2959 ACACAGGTATAGATTAAATTTCCACCACACACCATCATGATATTATTCAATTGAGGAC 3018  
QY 409 GGTACTGAGGTATTAATGTTCCCAATATGGATATATTATGTTTCTAGCACTGTTAAT 468  
Db 3019 TTAGCACAGCTCATTCATGATTTAAAAAATGCAATAGAGAGCTGATATTGCGATTAG 3078  
QY 469 AATGGAGCTGTAATTATGTTTACATCTGTAACCTCCAAATGCTACTATTTCATGAAAAGAAT 528  
Db 3079 CTTGTATCAAAAACCTGGCGTTGGAACATATAGCTTCAGGGGTAGCTAAAGCTTTCGCCGAT 3138  
QY 529 ACTGATGCGACATCGGGAGATGTTGGTGGAAAACTGTAGATCAAAAAAC 578  
Db 3139 AAAATTGTTATAGTGGTTGATGAGGTACAGGTGCATCGCCTAAAC 3188

RESULT 20  
US-08-119-125A-1  
; Sequence 1, Application US/08119125A  
; Patent No. 5610011  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, Hilda Elizabeth  
; APPLICANT: VECHT, Uri  
; TITLE OF INVENTION: DNA Sequences which code for Virulence  
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly  
; TITLE OF INVENTION: Antibodies derived therefrom and the use thereof for the diagn  
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Centraal Diergeneeskundig Instituut  
; STREET: Edelhertweg 15  
; CITY: PH Lelystad  
; STATE:  
; COUNTRY: The Netherlands  
; ZIP: NL-8219  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS v. 6.0  
; SOFTWARE: WordPerfect v. 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/119,125A  
; FILING DATE: 20-SEP-1993  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/NL92/00054  
; FILING DATE: 19-MAR-1992  
; APPLICATION NUMBER: NL 9100510  
; FILING DATE: 21-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Handal, Anthony H.  
; REGISTRATION NUMBER: 26275  
; REFERENCE/DOCKET NUMBER: SMITHHE119125  
; TELEPHONE: (203) 838-8589  
; TELEFAX: (203) 838-8794  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4376 base pairs  
; TYPE: Nucleic acid with corresponding amino acids  
; STRANDEDNESS: single stranded  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus suis type II (pathogenic)  
; FEATURE:  
; OTHER INFORMATION: Extracellular protein factor (EF) gene  
; NAME/KEY: promoter -35 region  
; LOCATION: bp 66 to 71  
; FEATURE:  
; NAME/KEY: promoter -10 region  
; LOCATION: bp 89 to 94  
; FEATURE:

; NAME/KEY: promoter -35 region  
; LOCATION: bp 153 to 158  
; FEATURE:  
; NAME/KEY: promoter -10 region  
; LOCATION: bp 176 to 181  
; FEATURE:  
; NAME/KEY: ribosome binding site  
; LOCATION: bp 350 to 356  
; FEATURE:  
; NAME/KEY: signal peptide  
; LOCATION: bp 361 to 498  
; FEATURE:  
; NAME/KEY: mature peptide  
; LOCATION: bp 499 to 2890  
; FEATURE:  
; NAME/KEY: dyad symmetry regions  
; LOCATION: from bp 4186 to 4198 and from bp 4203 to 4215  
; FEATURE:  
; NAME/KEY: dyad symmetry regions  
; LOCATION: from bp 4243 to 4257 and from bp 4263 to 4276  
; US-08-119-125A-1  
; Query Match 3.0%; Score 45.6; DB 1; Length 4376;  
; Best Local Similarity 51.5%; Pred. No. 0.072;  
; Matches 105; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 365 CTAATACTACACACAGTTCCACTGTTACTGAGTCAAATAATGATGCTACTGAGGTTATTA 424  
Db 3772 CTGTTATTAGCCAGGTACAGTTGTTACAGCATCTCTAAGACGAGGTTGGAAATAGTT 3831  
QY 425 ATGTTTCCCAATATGATATTAATTATGTTTCTAGCACTGTTTAATATGAGGCTGTAATTA 484  
Db 3832 CTGATTTCAGCAAGAGCTGAAGTTGTAGCAGTAGACGAAAAATAATTCTGCAGCAGGATGA 3891  
QY 485 TCGTTACATCTGTAACCTCAAATGCTACTATTTCATGAAAAGATACTGATGCGACATGG 544  
Db 3892 AAGTTAAATCAGTTACTACAAATGCTAATAATGTTGAGAAGAAAGCTAAGCAATTACCGA 3951  
QY 545 GAGATGGTGGTGGAAAACTGTAG 568  
Db 3952 ATACTGGTAGGAAGCAAAATTCAG 3975

RESULT 21  
US-08-119-125A-2  
; Sequence 2, Application US/08119125A  
; Patent No. 5610011  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, Hilda Elizabeth  
; APPLICANT: VECHT, Uri  
; TITLE OF INVENTION: DNA Sequences which code for Virulence  
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly  
; TITLE OF INVENTION: Antibodies derived therefrom and the use thereof for the diagn  
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Centraal Diergeneeskundig Instituut  
; STREET: Edelhertweg 15  
; CITY: PH Lelystad  
; STATE:  
; COUNTRY: The Netherlands  
; ZIP: NL-8219  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: MS-DOS v. 6.0  
; SOFTWARE: WordPerfect v. 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/119,125A  
; FILING DATE: 20-SEP-1993  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/NL92/00054  
; FILING DATE: 19-MAR-1992  
; APPLICATION NUMBER: NL 9100510  
; FILING DATE: 21-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Handal, Anthony H.  
; REGISTRATION NUMBER: 26275  
; REFERENCE/DOCKET NUMBER: SMITHHE119125  
; TELEPHONE: (203) 838-8589  
; TELEFAX: (203) 838-8794  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4376 base pairs  
; TYPE: Nucleic acid with corresponding amino acids  
; STRANDEDNESS: single stranded  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus suis type II (pathogenic)  
; FEATURE:  
; OTHER INFORMATION: Extracellular protein factor (EF) gene  
; NAME/KEY: promoter -35 region  
; LOCATION: bp 66 to 71  
; FEATURE:  
; NAME/KEY: promoter -10 region  
; LOCATION: bp 89 to 94  
; FEATURE:

```

> FILING DATE: 19-MAR-1992
> APPLICATION NUMBER: NL 9100510
> FILING DATE: 21-MAR-1991
> ATTORNEY/AGENT INFORMATION:
> NAME: Handal, Anthony H.
> REGISTRATION NUMBER: 26275
> REFERENCE/DOCKET NUMBER: SMITHHE119125
> TELECOMMUNICATION INFORMATION:
> TELEPHONE: (203) 838-8589
> TELEFAX: (203) 838-8794
> INFORMATION FOR SEQ ID NO: 2:
> SEQUENCE CHARACTERISTICS:
> LENGTH: 6744 base pairs
> TYPE: Nucleic acid with corresponding amino acids
> STRANDEDNESS: single stranded
> TOPOLOGY: linear
> MOLECULE TYPE: genomic DNA
> ORIGINAL SOURCE:
> ORGANISM: Streptococcus suis type II (pathogenic)
> FEATURE:
> OTHER INFORMATION: Extracellular factor related protein (EF*) gene
> FEATURE:
> NAME/KEY: promoter -35 region
> LOCATION: bp 66 to 71
> FEATURE:
> NAME/KEY: promoter -10 region
> LOCATION: bp 89 to 94
> FEATURE:
> NAME/KEY: promoter -35 region
> LOCATION: bp 153 to 158
> FEATURE:
> NAME/KEY: promoter -10 region
> LOCATION: bp 176 to 181
> FEATURE:
> NAME/KEY: ribosome binding site
> LOCATION: bp 350 to 356
> FEATURE:
> NAME/KEY: signal peptide
> LOCATION: bp 361 to 498
> FEATURE:
> NAME/KEY: start of repetitive units R1-R11
> LOCATION: bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,
> LOCATION: 5065, 5293, 5521:
> FEATURE:
> NAME/KEY: start of repetitive Asn-Pro-Asn-Leu sequences
> LOCATION: bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,
> LOCATION: 5128, 5356, 5584:
> FEATURE:
> NAME/KEY: dyad symmetry regions
> LOCATION: from bp 6554 to 6566 and from bp 6571 to 6583
> FEATURE:
> NAME/KEY: dyad symmetry regions
> LOCATION: from bp 6611 to 6625 and from bp 6631 to 6644
>
> US-08-119-125A-2
>
> Query Match 3.0%; Score 45.6; DB 1; Length 6744;
> Best Local Similarity 51.5%; Pred. No. 0.084;
> Matches 105; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
>
> Qy 365 CTAATACTACACAGTTTCCACTCTTACTGAGTCAAATAATGATGGTACTGAGGTTATTA 424
> Db 6140 CTGTTATTAAAGCCAGGTACAGTCTGTACAGCATACTCTAAAGACGACGAGTTGGAAATAGTT 6199
>
> Qy 425 ATGTTTCCCAATATGGATATATTATGTTTCTAGACACTGTTTAATAATGGAGCTGTAATTA 484
> Db 6200 CTGATTGACGAGAAGCTGAAAGTTGTAGCAGTAGACGAAATAATTTCTGCAGCAGGAGTGA 6259
>
> Qy 485 TGGTTACATCTGTAACTCCAAATGCTACTATTTCATGAAAGAAATACATGATCGCAGCATGGG 544
> Db 6260 AAGTTAATCAGTTTACTACAAATGCTTAATATGTTTGAGAGAAAGCTTAAGCAATTACCGA 6319
>
> Qy 545 GAGATGTGGTGGAAAACTGTAG 568
>

```

TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-1377

Query Match 3.0%; Score 44.6; DB 4; Length 11679;  
Best Local Similarity 44.4%; Pred. No. 0.18;  
Matches 234; Conservative 0; Mismatches 284; Indels 9; Gaps 1;

QY 42 AGCATTTGGTATGGCTGATCAACAGTTAGCGGATAGCTTTTGGCGTGACAGGAC 101  
DB 663 AGCATCAGGTAATGTCCACATTACAGCTGTTTGAATAAACAATCCAGCAGATGCGGCCAA 722  
QY 102 AATTACAGTTCAAGATACCTCAAAAGCGCAACTATAAAGATATAAAGATTTTGTATGC 161  
DB 723 CACGGCGTTACAGTTGTTATATGCGGTAACTTATATGCGCTGTAGATAAAGCAGC 782  
QY 162 AGAAATAGATAAATGCAAAATGATCTGATTCGAATAAAGATGAGCTTCTTATTTAATCC 221  
DB 783 AGGTACATGACAGTAAGTGTACCGGCGAGTGGCTTGTAGTTGGGATGCAGATAAGACGAT 842  
QY 222 TCAAGGTAAAGACGTGAGTATAAAGCTTCAACTGATTTTAACTCTTTTACGACAC 281  
DB 843 TGAATGCCAAAGTAACGTTTACAGATGCGAGGTAAACAGCAGTGTTAATGATACACA 902  
QY 282 TACTAATGAGGAGAGACATATGTAACCTAAAGAGATCTGCGTCAGCAAAATGAGATTGC 341  
DB 903 AACTTAT-----ACACTTGATACCACTGCACCAATGCACCACTGATTGACCCAGT 953  
QY 342 GACATGGGCTAAATCTATATCAGCTAATACTACACCAAGTTTCCAGTGTACTGAGTCAAA 401  
DB 954 TAACGGGACAGACCCGATTACAGGTACAGCAGAACCTGGTTCAACAGTGTGACTTGA 1013  
QY 402 TAATGATGTTACTGAGCTTATTAATGTTTCCCAATATGATATTTATGTTTCTAGCAC 461  
DB 1014 TCCGAGGTGAGTACGAAAACCTGTTGAGCAGCAGCGATGTTGACTGTGCGCAAA 1073  
QY 462 TGTTAATAATGAGCTGTAATATGTTTACATCTGTAACCTCCAAATGCTACTAATCATGA 521  
DB 1074 CCCAGGTTTGAATGATGGCGATGAAGTTTACAGCTGTAGCAACAGACCCCTGCTGCAATAC 1133  
QY 522 AAGAATACTGATGCGCATGCGGAGATGTTGGTGGAAAACCTGAG 568  
DB 1134 ATCAGGCCCAACGACGAGTGGTTGATGCTGTTGCGCCGACAGTAG 1180

RESULT 24  
US-09-601-198-166  
; Sequence 166, Application US/09601198  
; Patent No. 6531583  
; GENERAL INFORMATION:  
; APPLICANT: Cassell, Gail H.  
; APPLICANT: Chen, Ellison Y.  
; APPLICANT: Glass, Jennifer S.  
; APPLICANT: Glass, John I.  
; APPLICANT: Heiner, Cheryl R.  
; APPLICANT: Lefkowitz, Elliot  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA  
; FILE REFERENCE: UREALYTICUM  
; CURRENT APPLICATION NUMBER: US/09/601,198  
; CURRENT FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/073,189  
; PRIOR FILING DATE: 1998-01-30  
; NUMBER OF SEQ ID NOS: 181  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 166  
; LENGTH: 606  
; TYPE: DNA  
; ORGANISM: Ureaplasma urealyticum  
US-09-601-198-166

Query Match 2.9%; Score 44.2; DB 4; Length 606;  
Best Local Similarity 44.3%; Pred. No. 0.081;

Matches 181; Conservative 0; Mismatches 228; Indels 0; Gaps 0;

QY 123 AAAAGGCGCAACCTATAAGCATATAAAGTTTGTGCGAGAAATAGATAATGCAAAATGT 182  
DB 63 AAGATATGAATTCAAAACGAAAAAACAATTCGAATAAATAATTAAGTTGTGAATGA 122  
QY 183 ATCTGATTTCGAATAAAGATGGAGCTTCTTAATTTAAATTCCTCAAGGTAAAGAGCTGAGTA 242  
DB 123 TGTGATATTAGTACCTTTATATGATTTTAAAGATAAAAAATCTAACTATTATATACGAAAAC 182  
QY 243 TAAAGCTTCAACTGATTTAATTTCTTTTACGACAACTACTAATGGAGGAGAACATA 302  
DB 183 ATTTTATTTCCATAATATCAATTTGCAATGCTAAATAAAAAAATTTATTTTATTTCTAA 242  
QY 303 TGTAACTAAAAAGATACCTGCGTCAGCAAAATGAGATTGCGACATGGGCTAAATCTATATC 362  
DB 243 AAAAATCTAATTAATCTGTTTATACAGAAAAATTTATAGATATAGGATTTTACTTGAACAAT 302  
QY 363 AGCTAATACTACACAGTTTCCACTGTTTACGTGAGTCAAAATGAATGATGGTACTGAGGTAT 422  
DB 303 TATTTATTAATGATCAAAACAAAACTTATATCAGGCCAACTGATTTTATGACCAAAAAATTT 362  
QY 423 TAAATGTTTCCCAATATGATATTTATGTTTCTAGCACTGTTAATAATGGAGCTGTAAT 482  
DB 363 TATTTGATTTAAAAAATGAAATTAATAATGCTACTCTGTTTCGAAAAATTTTATATGATAT 422  
QY 483 TATGTTTACATCTGTAACTCCAAATGCTACTATTTATCATGAAAGAATACT 531  
DB 423 GATTTGTTAAAAATGAATCAATTAATAATTAATCAATAGTGAATGAATCT 471

RESULT 25  
US-09-134-001C-322  
; Sequence 322, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 322  
; LENGTH: 30549  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-322

Query Match 2.9%; Score 44.2; DB 4; Length 30549;  
Best Local Similarity 45.2%; Pred. No. 0.32;  
Matches 163; Conservative 0; Mismatches 198; Indels 0; Gaps 0;

QY 180 TGTATCTGATTTCGAATAAAGATGGAGCTTCTTATTTAATTCCTCAAGGTAAAGAGCTGA 239  
DB 5826 TATTCCTAATAACAAATACAACTATAGATTCTAATGTTATAGCTACTGTAAACAATTC 5885  
QY 240 GTATAAAGCTTCAACTGATTTTAAATTTCTTTTACGACAACTACTAATGGAGGGAGAAC 299  
DB 5886 CACTCTACCAACCGGAAATATTACTGTCTAAAACCTCAATGACAAATATGTAACGTACAC 5945  
QY 300 ATATGTAACTAAAAAGATACCTGCGTCAGCAAAATGAGATTGGACATGGGCTAAATCTAT 359  
DB 5946 TAAACAAATAGTAGTGGAAATGCTTCAAAATACAACTGAAGATATAAGTGTTTTTCAGA 6005  
QY 360 ATCAGCTAATACTACACAGGTTTCCACTGTTTACTGATCBAATATGATGGTACTGAGT 419  
DB 6006 AAACAGTGATCAAGTAAATGTTTACCCTGGCATGCAAGCTAAAAATGATGGTATTAAAT 6065

**Qy** 420 TATTAAATGTTCCCAATAAGTATTAATGTCTTAGCAGCTGTTAATAATGGAGCTGT 479  
| | | | |  
**Db** 6066 AATTAAAGGTCAAACTAATAATTTTAATGACTTCATATAGTTTTCATAGTAATATACCAGC 6125  
| | | | |  
**Qy** 480 AATTATGGTTACATCTGTAACCTCCAATGCTACTATTTCATGAAGAAGAACTATGATCGCAC 539  
| | | | |  
**Db** 6126 CCATTCTACTCTTACATGGAAGGAGGCCCTAATAGTTGGAAAAACAACATCGGTACTAC 6185  
| | | | |  
**Qy** 540 A 540  
|  
**Db** 6186 A 6186

**RESULT 26**  
US-09-134-001C-1028  
; Sequence 1028, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 1028  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1028

Query Match 2.9%; Score 44; DB 4; Length 1716;  
Best Local Similarity 42.9%; Pred. No. 0.13;  
Matches 218; Conservative 0; Mismatches 290; Indels 0; Gaps 0;

**Qy** 156 TGATGCAGAAATAGATATGCAAAATGTATCTGATTTCGAAATAAGATGGAGCTCTTTATT 215  
| | | | |  
**Db** 1131 TGAAGAAAACAATCAAACCTGCACCTGGTAACTGGTAACTGCACTGGAAATCTAAATCA 1190  
| | | | |  
**Qy** 216 AATTCTCAAGGTAAAGAAGCTGAGTATAAAGCTTCAACTGATTTTAAATTTCTTTTTAC 275  
| | | | |  
**Db** 1191 AGACACACAAGTAAACCTAATACAGATCAGACTGCTACTGGTACTCCAGCTGGTACTGA 1250  
| | | | |  
**Qy** 276 GACAACCTAATGAGGGAGAACATATGTAACTAAAGAAATGCTGGTGAGCAAAATGA 335  
| | | | |  
**Db** 1251 TAATCAAAAATACGCAAAACAAAGCAATCTGAAACAAAATAACCAAAATGCTCAACCATCAGC 1310  
| | | | |  
**Qy** 336 GATTGCGACATGGGTAAATCTATACAGTAACTACTACACCGATTTCCACTGTGTACTGA 395  
| | | | |  
**Db** 1311 ACCTGGACAACTGATCAAGCCGGTGCTACGGTTAACTGGATGCTGCCAAATCAAGA 1370  
| | | | |  
**Qy** 396 GTCAAAATAATGATGCTAGGAGTTTAAATGTTTCCCAATATGGATATTAATGTTTC 455  
| | | | |  
**Db** 1371 CGCTGAAGCAACCCCTAATCCAGATCAAAATAATACCAACTAATGTCGAAGCGGGA 1430  
| | | | |  
**Qy** 456 TAGCAGCTGTTAATTAATGGAGCTGTAATTAATGGTTACATCTGTAACTCCAAATGCTACTAT 515  
| | | | |  
**Db** 1431 TCAAAACAAATCAAAGTACACAAGATGATAATGATAATCAAAATACGCAACGAAGGTAAATAC 1490  
| | | | |  
**Qy** 516 TCATGAAAGAAATAGTATGCGACATGGGGAGATGGTGGAAAAAATCTAGATCAAAA 575  
| | | | |  
**Db** 1491 TAAAACAAAACAAATCAAAATGCTGAACAAAGGTAATCTGCGGTACAGATAAAGATGCTAA 1550  
| | | | |  
**Qy** 576 AACGTACTCGGTTGGTGATACAGTCAAAATATATCTATTACTTTAAGAATGCAGTCAATTA 635  
| | | | |  
**Db** 1551 AGTAATACTAATGATGATGAAACAAAACAGATGTTAAAAATGATCTGAATTAACCTGA 1610  
| | | | |  
**Qy** 636 TCATGGGTACAGAAAAGGTATCAATAT 663  
| | | | |

## RESULT 28

US-09-889-572-3  
; Sequence 3, Application US/09889572  
; Patent No. 6610517  
; GENERAL INFORMATION:  
; APPLICANT: Lubitz, Werner  
; TITLE OF INVENTION: COMPARTMENTALIZATION OF RECOMBINANT POLYPEPTIDES IN HOST CELLS  
; FILE REFERENCE: 100564-00070  
; CURRENT APPLICATION NUMBER: US/09/889,572  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: PCT/EP00/00686  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: DE 199 03 345.5  
; PRIOR FILING DATE: 1999-01-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 2766  
; TYPE: DNA  
; ORGANISM: Bacillus stearothermophilus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2763)  
; OTHER INFORMATION:  
; NAME/KEY: sig\_peptide  
; LOCATION: (1)..(93)  
; OTHER INFORMATION:  
; NAME/KEY: mat\_peptide  
; LOCATION: (94)..(2763)  
; OTHER INFORMATION:  
US-09-889-572-3

Query Match 2.9%; Score 43.4; DB 4; Length 2766;

Best Local Similarity 42.9%; Pred. No. 0.22;  
Matches 215; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY	29	TAGTGGCGAGTTTAGCATTGGTATGGCTGTATCCAGTTACCGGATAGCTTTTGGCG	88
Db	1022	TAGTTCAAGTGGGAAGTATCAGTATCTGCTCTAAAGTTACAGCTGTAAACCGGAA	1081
QY	89	CTGAGACAGGACAAATTPACAGTTCAAGTACTCAAAAGCGCAACTATATAAGCATATA	148
Db	1082	CAGCTGATGTTACTGCAAAAGTTACATTACCAGATGGTGTGTACTAAACAAATACATTTA	1141
QY	149	AAGTTTTGATGCAGAAATAGATAATGCAATGTATCTGATTGCAATAAGATGGAGCTT	208
Db	1142	AAGTGACAGTTACAGAAGTGCCTGTGCAAGTACAAATCAAGGATTTACTTTAGTTGATA	1201
QY	209	CTTATTTAAATTCCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTC	268
Db	1202	ATCTTTCTAATGCTCCACAGATACAGTTGCAITTTAACAAGCTGAGAAAGTAACTTCAA	1261
QY	269	TTTTTACGACAACTACTAATGGAGGAGACATATGTAATCACTAAAGATAGTGGCTCAG	328
Db	1262	TGTTTGTGGAGAACTAAACACAGTTGCAATGTATGATATCTAAACACGGTGATCCTGAA	1321
QY	329	CARATGAGATTGGACATGGCTAAATCTATATCAGCTAATCTACACCAAGTTTCCACTG	388
Db	1322	CTAAACCTGTTGATTTCAAGATGCACTGTACGTTCAITTAATCCATTTATTTGCAACAG	1381
QY	389	TTACTGAGTCAAAATATGATGGTACTGAGGTTAATAATGTTTCCCAATATGGATATATT	448
Db	1382	CTGCTATTATATGGTAGTGAGCTCTTGTACAGCTAATGCTGGCCAATCTGGAAAAGCTT	1441
QY	449	ATGTTTCTAGACCTGTTAAATAGTAAATTTATGTTTGTATCTGTAACCTCCAATG	508
Db	1442	CATTGGAAGTAAACATTTAAAGATAATACAAAAGACATTTACAGTTGATGTGAAAAG	1501
QY	509	CTACTATTTCATGAAAAGATA	529
Db	1502	ACCTGTATTACAGATATTA	1522

## RESULT 30

US-08-257-073-10  
; Sequence 10, Application US/08257073

## RESULT 29

US-09-463-402-13  
; Sequence 13, Application US/09463402  
; Patent No. 6596510  
; GENERAL INFORMATION:  
; APPLICANT: Lubitz, Werner  
; APPLICANT: Resch, Stephanie  
; TITLE OF INVENTION: Secretion of Carrier-bound proteins into the Periplasm  
; TITLE OF INVENTION: and into the Extracellular Space  
; FILE REFERENCE: 05649059  
; CURRENT APPLICATION NUMBER: US/09/463,402  
; CURRENT FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: DE19732829.6  
; PRIOR FILING DATE: 1997-07-30  
; PRIOR APPLICATION NUMBER: PCT/EP98/04723  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 4065  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: fusion of SbeB  
; OTHER INFORMATION: with Male  
US-09-463-402-13

Query Match 2.9%; Score 43.4; DB 4; Length 4065;

Best Local Similarity 42.9%; Pred. No. 0.25;  
Matches 215; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

QY	29	TAGTGGCGAGTTTAGCATTGGTATGGCTGTATCCAGTTACCGGATAGCTTTTGGCG	88
Db	2180	TAGTTTCAAGTGGGGAAGTATCAGTATCTGCTGTCTAAAGTTACAGCTGTAAACCGGAA	2239
QY	89	CTGAGACAGGACAAATACAGTTCAAGTACTCAAAAGGCGCACTATATAAGCATATA	148
Db	2240	CAGCTGATGTTACTGCAAAAGTTACATTACCAGATGGTGTGTACTAAACAAATACATTTA	2299
QY	149	AAGTTTTTGTGTCGCAAAATAGATAATGCAATGTATCTGATTTCGAATAAAGATGGAGCTT	208
Db	2300	AAGTGACAGTTACAGAAGTGCCTGTGCAAGTACAAAATCAAGGATTTACTTTAGTTGATA	2359
QY	209	CTTATTTAAATTCCTCAAGGTAAAGAGCTGAGTATAAAGCTTCAACTGATTTTAAATTC	268
Db	2360	ATCTTTCTAATGCTCCACAGAAATACAGTTGCAITTTAACAAGCTGAGAAAGTAACTTCAA	2419
QY	269	TTTTTACGACAACTACTAATGGAGGAGACATATGTAACATAAAGATAGTGGCTCAG	328
Db	2420	TGTTTGTGGAGAACTAAACACAGTTGCAATGTATGATATCTAAACACGGTATCCTGAA	2479
QY	329	CAAAATGAGATTGCGACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACTG	388
Db	2480	CTAAACCTGTTGATTTCAAGATGCACTGTACGTTTCAITTAATCCATTTATTTGCAACAG	2539
QY	389	TTACTGAGTCAAAATATGATGGTACTGAGGTTAATAATGTTTCCCAATATGGATATATT	448
Db	2540	CTGCTATTAAATGGTAGTGAGCTCTTGTGTCAGAGCTAATGCTGGCCAATCTGAAAAGCTT	2599
QY	449	ATGTTTCTAGCAGCTGTTAATAATGGAGCTGTAATTTATGTTTACATCTGTAACCTCAAAATG	508
Db	2600	CATTGGAAGTAAACATTTAAAGATAATACAAAAGACATTTACAGTTGATGTGAAAAGAG	2659
QY	509	CTACTATTTCATGAAAAGATA	529
Db	2660	ACCTGTATTACAGATATTA	2680

## RESULT 30

US-08-257-073-10  
; Sequence 10, Application US/08257073

Patent No. 5766597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: de Taisne, Charles  
APPLICANT: Tine, John A.  
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,073  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/075,783  
FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,183  
FILING DATE: 20-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2570  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5181 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-073-10

Query Match 2.9%; Score 43.2; DB 1; Length 5181;  
Best Local Similarity 44.4%; Pred.No. 0.3;  
Matches 220; Conservative 0; Mismatches 273; Indels 3; Gaps 1;  
QY 158 ATGCAGAAATAGATAATGCAAAATGTATCTGATTGCAATAAGATGGAGCTTCTTATTAA 217  
Db 1130 ACGAAGAAAAATAAAGAAATGGCAAACTATTAAATTTACATTGATGATTATTTA 1189  
QY 218 TTCTCAAGGTAAAGAGCTGAGTATATAAGCTTCAACTGATTTAAATTCCTTTTACGA 277  
Db 1190 CTGATCCACTTCAATTAGATAATTATTTAAGAGAAAAATAAAAAAGTTGATGTAAACAC 1249  
QY 278 CAACTACTAATGGAGGAGACATATGTAACTAAAAAGATACTCGCTCAGCAATGAGA 337  
Db 1250 CTAAATCAAGATCCTACGAAATCTGTTCAAAATACCAAAAGTTCTTTATCCAAATGGTA 1309  
QY 338 TTGCGACATGGGCTAAATCTATATCAGCTTAATACACACAGTTTCCACTGTACTGAGT 397  
Db 1310 TTG---TATATCTTTACCACCTCATGATTTCAATTAATTCATTAGCTGCAGATAATGATA 1366  
QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGGATATTATTATGTTCTTA 457  
Db 1367 AAAATTCATATGGTATGATTTAATGAATCCTGATCTAAAGAAAAAATTAATGAAAAATTA 1426  
QY 458 GCACGTGTTAATAGGAGCTGTAATTATGGTTACATCTGTAACCTCCAAATGCTACTATTC 517

Db 1427 TTACAGATAATAAGGAAAGAAAAATATTCTAATAACATTAAAAAACAATGATTAG 1486  
QY 518 ATGAAAAAGAAATACCTGATCGCAGCATGGGAGATGGTGGGAAAAAATCTGTAGATCAAAAAA 577  
Db 1487 AAGAAAAAATAATTAATCACACAAAAAGAACAAAAATTAATTTGAAAGATTATGAA 1546  
QY 578 CGTACTCGGTTGGTGATACAGTCAAAATATATCTATTACTTTATAAGAAATGCAATCAATTATC 637  
Db 1547 AGTCAAAAAAGGATTATGAAGATTACTTGAATAATTTTATGAAATGAAATTTAATAATA 1606  
QY 638 ATGCTACAGAAAAAGT 653  
Db 1607 ATTTGACAAAGATGT 1622

Search completed: July 27, 2004, 01:16:20  
Job time : 130 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 00:04:34 ; Search time 727 Seconds  
(without alignments)  
10146.415 Million cell updates/sec

Title: US-10-009-254-1

Perfect score: 1509

Sequence: 1 atgaaaagaataatgattca.....gtcgtcgttaccttcttaa 1509

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3216467 seqs, 2444149694 residues

Total number of hits satisfying chosen parameters: 6432934

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1509	100.0	1509	12	US-10-333-002-15
2	64.8	4.3	1983	13	US-10-206-576-217
3	64.8	4.3	2199	12	US-10-333-002-28
4	64.8	4.3	15614	9	US-09-070-927A-45
5	63.2	4.2	1881	17	US-10-661-809-12
6	59.2	3.9	1137	13	US-10-282-122A-15988
7	53	3.5	4985	15	US-10-094-340-10
8	53	3.5	4985	15	US-10-056-405-10
9	50.4	3.3	14066	15	US-10-349-680-149
10	50.4	3.3	14067	13	US-10-282-122A-40681
11	50	3.3	3996	15	US-10-087-464-42
12	49.8	3.3	1687	13	US-10-206-576-219
13	48.6	3.2	4997	13	US-10-282-122A-35506
14	48.4	3.2	3666	9	US-09-137-531-13

48.4	3.2	3666	9	US-09-137-531-14	Sequence 14, Appl
48.4	3.2	4197	9	US-09-137-531-7	Sequence 7, Appl
48.4	3.2	4197	9	US-09-137-531-8	Sequence 8, Appl
47.4	3.1	1037	16	US-10-373-667-3	Sequence 3, Appl
47.4	3.1	1159	16	US-10-373-667-1	Sequence 1, Appl
47.4	3.1	1471	16	US-10-373-667-2	Sequence 2, Appl
47.4	3.1	4047	9	US-09-815-242-4843	Sequence 4843, Ap
47.4	3.1	4050	9	US-09-815-242-9039	Sequence 9039, Ap
47.4	3.1	4358	8	US-08-781-986A-454	Sequence 454, App
47.4	3.1	4358	13	US-10-329-624-454	Sequence 454, App
47.2	3.1	5314	15	US-10-155-533-1	Sequence 1, Appli
46.8	3.1	2801	13	US-10-282-122A-34844	Sequence 34844, A
46.8	3.1	7047	15	US-10-240-453-260	Sequence 260, App
46.2	3.1	4158	13	US-10-282-122A-8140	Sequence 8140, Ap
45.8	3.0	684707	16	US-10-398-221-9	Sequence 9, Appli
45.8	3.0	3011208	16	US-10-398-221-2058	Sequence 2058, Ap
45.6	3.0	2017	15	US-10-155-533-3	Sequence 3, Appli
45.6	3.0	6397	13	US-10-221-714A-181	Sequence 181, App
45.6	3.0	6397	15	US-10-239-676-107	Sequence 107, App
45.6	3.0	6397	15	US-10-311-455-1319	Sequence 1319, Ap
45.6	3.0	6397	15	US-10-240-453-119	Sequence 119, App
45.4	3.0	2217	13	US-10-282-122A-17439	Sequence 17439, A
45.2	3.0	3931	15	US-10-006-780-1	Sequence 1, Appli
45	3.0	12237	15	US-10-311-455-2331	Sequence 2331, Ap
44.8	3.0	573	15	US-10-029-386-25433	Sequence 25433, A
44.8	3.0	5358	17	US-10-637-544-19	Sequence 19, Appl
44.8	3.0	16287	15	US-10-311-455-645	Sequence 645, App
44.4	2.9	7446	13	US-10-282-122A-7578	Sequence 7578, Ap
44.2	2.9	486	13	US-10-465-217-13	Sequence 13, Appl
44.2	2.9	606	15	US-10-349-680-85	Sequence 85, Appl
44	2.9	3673778	15	US-10-313-841-1	Sequence 1, Appli
43.8	2.9	528	15	US-10-029-386-5350	Sequence 5350, Ap
43.8	2.9	807	13	US-10-282-122A-16974	Sequence 16974, A
43.8	2.9	1243	13	US-10-282-122A-9076	Sequence 9076, Ap
43.8	2.9	1910	15	US-10-155-533-2	Sequence 2, Appli
43.8	2.9	3532	15	US-10-017-161-1913	Sequence 1913, Ap
43.8	2.9	3532	16	US-10-292-798-1569	Sequence 1569, Ap
43.6	2.9	1590	15	US-10-032-585-6364	Sequence 6364, Ap
43.4	2.9	2000	9	US-09-938-842A-3954	Sequence 3954, Ap
43.4	2.9	2000	11	US-09-938-842A-3954	Sequence 3954, Ap
43.2	2.9	475	9	US-09-864-761-1619	Sequence 1619, Ap
43	2.8	483	16	US-10-398-221-365	Sequence 365, App
43	2.8	483	16	US-10-398-221-2558	Sequence 2558, Ap
43	2.8	2001	9	US-09-737-178-84	Sequence 84, Appl
43	2.8	2001	10	US-09-853-079-84	Sequence 84, Appl
43	2.8	2001	17	US-10-294-443-84	Sequence 84, Appl
43	2.8	3057	15	US-10-349-680-148	Sequence 148, App
43	2.8	3402	9	US-09-737-178-86	Sequence 86, Appl
43	2.8	3402	10	US-09-853-079-86	Sequence 86, Appl
43	2.8	3402	17	US-10-294-443-86	Sequence 86, Appl
43	2.8	8201	15	US-10-363-798-1	Sequence 1, Appli
42.8	2.8	1959	9	US-09-864-761-4012	Sequence 4012, Ap
42.8	2.8	2058	13	US-10-282-122A-7579	Sequence 7579, Ap
42.8	2.8	2769	15	US-10-032-585-6675	Sequence 6675, Ap
42.6	2.8	663	13	US-10-027-632-250796	Sequence 250796,
42.6	2.8	663	16	US-10-027-632-250796	Sequence 250796,
42.6	2.8	663	16	US-10-027-632-250797	Sequence 250797,
42.6	2.8	2247	13	US-10-424-559-4419	Sequence 4419, A
42.6	2.8	168575	15	US-10-178-194-1	Sequence 1, Appli
42.4	2.8	1224	13	US-10-282-122A-16026	Sequence 16026, A
42.4	2.8	1495	16	US-10-398-221-3257	Sequence 3257, Ap
42.4	2.8	2766	9	US-09-117-447-5	Sequence 5, Appli
42.4	2.8	5917	15	US-10-087-464-9	Sequence 9, Appli
42.2	2.8	810	15	US-10-349-680-126	Sequence 126, App
42.2	2.8	3426	13	US-10-282-122A-34067	Sequence 34067, A
42.2	2.8	4185	15	US-10-304-095-7	Sequence 7, Appli
42.2	2.8	10640	15	US-10-304-095-5	Sequence 5, Appli
42.2	2.8	58985	10	US-09-901-152-3	Sequence 3, Appli
42.2	2.8	136601	10	US-09-855-824-3	Sequence 3, Appli
42	2.8	5379	15	US-10-311-455-1650	Sequence 1650, App
42	2.8	5379	15	US-10-240-453-130	Sequence 130, App
42	2.8	6060	8	US-08-781-986A-534	Sequence 534, App

C	88	42	2.8	6060	13	US-10-329-624-534	Sequence 534, App
	89	42	2.8	7104	9	US-09-815-242-4580	Sequence 4580, Ap
	90	42	2.8	7107	9	US-09-815-242-8291	Sequence 8291, Ap

ALIGNMENTS

RESULT 1

US-10-333-002-15

;

Sequence 15, Application US/10333002

;

Publication No. US20040071729A1

;

GENERAL INFORMATION:

;

APPLICANT: Adderson, Elisabeth

;

APPLICANT: Bohnsack, John

;

TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic

;

FILE OF INVENTION: Compositions and Vaccines Thereof

;

FILE REFERENCE: 2511-1-001 (SJ-0039)

;

CURRENT APPLICATION NUMBER: US/10/333,002

;

CURRENT FILING DATE: 2003-07-08

;

PRIOR APPLICATION NUMBER: PCT/US01/24795

;

PRIOR FILING DATE: 2001-08-08

;

PRIOR APPLICATION NUMBER: US 09/634,341

;

PRIOR FILING DATE: 2000-08-08

;

NUMBER OF SEQ ID NOS: 37

;

SOFTWARE: PatentIn version 3.1

;

SEQ ID NO 15

;

LENGTH: 1509

;

TYPE: DNA

;

ORGANISM: Streptococcus agalactiae

US-10-333-002-15

Query Match

Best Local Similarity 100.0%; Score 1509; DB 12; Length 1509;

Matches 1509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1

ATGAAAAAGAAATGATTC

CAATCGCTGTTAGTGGCGAGTTT

AGCAATTTGGTATGGCTGTA

60

DB

1

ATGAAAAAGAAATGATTC

CAATCGCTGTTAGTGGCGAGTTT

AGCAATTTGGTATGGCTGTA

60

QY

61

TCACCAAGTTACGCGATAGCTTT

TGCGCTGAGACAGGACAAAT

TACAGTTCAGATATCT

120

DB

61

TCACCAAGTTACGCGATAGCTTT

TGCGCTGAGACAGGACAAAT

TACAGTTCAGATATCT

120

QY

121

CAAAAAGGCGCAACCTATAA

AGCATATAAAGTTTTGATG

CAGAAATAGATAATGCAAA

180

DB

121

CAAAAAGGCGCAACCTATAA

AGCATATAAAGTTTTGATG

CAGAAATAGATAATGCAAA

180

QY

181

GTATCTGATTGCAATAAAGAT

TGGAGCTTCTTATTTAATTC

CTCAAGGTAAGAAGCTGAG

240

DB

181

GTATCTGATTGCAATAAAGAT

TGGAGCTTCTTATTTAATTC

CTCAAGGTAAGAAGCTGAG

240

QY

241

TATAAAGCTTCAACTGATTTT

AAATCTCTTTTACGCAACT

ACTAATAATGGAGGAGACA

300

DB

241

TATAAAGCTTCAACTGATTTT

AAATCTCTTTTACGCAACT

ACTAATAATGGAGGAGACA

300

QY

301

TATGTAACATAAAAAGATAC

TGGCTCAGCAAAATGAGATT

GCGACATGGGCTAAATCTATA

360

DB

301

TATGTAACATAAAAAGATAC

TGGCTCAGCAAAATGAGATT

GCGACATGGGCTAAATCTATA

360

QY

361

TCAGCTAAATACACAGTTTCC

ACTGTTACGTAAGTCAAAAT

ATGATGGTACTGAGGTT

420

DB

361

TCAGCTAAATACACAGTTTCC

ACTGTTACGTAAGTCAAAAT

ATGATGGTACTGAGGTT

420

QY

421

ATTAATGTTTTCCCAATATGG

ATATTAATTTTCTAGCACTG

TTAATAATGGAGCTGTA

480

DB

421

ATTAATGTTTTCCCAATATGG

ATATTAATTTTCTAGCACTG

TTAATAATGGAGCTGTA

480

QY

481

ATTATGGTTACATCTGTAATC

CCAAATGCTACTATTATGAA

AGAAATACTGATCCGACA

540

DB

481

ATTATGGTTACATCTGTAATC

CCAAATGCTACTATTATGAA

AGAAATACTGATCCGACA

540

QY

541

TGGGGAGATGGTGGTGGAAA

AACTAGATCAAAAAACGTA

CTCGGTTGGTGATACAGTC

600

DB	541	TGGGGAGATGGTGGTGGAAA	AACTAGATCAAAAAACGTA	CTCGGTTGGTGATACAGTC	600				
QY	601	AAATATACATTAATTAAGAA	TGCAGTCAATTAATCATG	TACAGAAAAAGTGTATCAA	660				
DB	601	AAATATACATTAATTAAGAA	TGCAGTCAATTAATCATG	TACAGAAAAAGTGTATCAA	660				
QY	661	TATGTTATAAAGGATACAT	TGCCATCTCTCTGCTAG	TTGATTGAAACGAAGGTC	720				
DB	661	TATGTTATAAAGGATACAT	TGCCATCTCTCTGCTAG	TTGATTGAAACGAAGGTC	720				
QY	721	GAAATTAACATTAATTAAG	GAATTAATTAACACTCTA	CAAGGTTTCGAAAAA	780				
DB	721	GAAATTAACATTAATTAAG	GAATTAATTAACACTCTA	CAAGGTTTCGAAAAA	780				
QY	781	GCAACTGGGAAGTATAAC	CTGTAGAGGAAATAATA	TTTCAAGATTAATTC	840				
DB	781	GCAACTGGGAAGTATAAC	CTGTAGAGGAAATAATA	TTTCAAGATTAATTC	840				
QY	841	GCAGCTACCAATACATC	CAACCGGAAATTAAT	TAATTAATTAATTAAT	900				
DB	841	GCAGCTACCAATACATC	CAACCGGAAATTAAT	TAATTAATTAATTAAT	900				
QY	901	AAGGGAATAAATACAA	TCACTCAAGGATTAAT	TAATTAATTAATTAAT	960				
DB	901	AAGGGAATAAATACAA	TCACTCAAGGATTAAT	TAATTAATTAATTAAT	960				
QY	961	GGTTCAAGCTGATTAAC	CAAGAAATTAATTAAT	TAATTAATTAATTAAT	1020				
DB	961	GGTTCAAGCTGATTAAC	CAAGAAATTAATTAAT	TAATTAATTAATTAAT	1020				
QY	1021	GATGACCCAGGTCAAAA	AGTAAACAGTGGGAT	GGTCAAAATTAATTAAT	1080				
DB	1021	GATGACCCAGGTCAAAA	AGTAAACAGTGGGAT	GGTCAAAATTAATTAAT	1080				
QY	1081	GGTTCCCAAAAAGCTTC	ATTAAGGTCATATTTG	TTTAAAGATGCTAC	1140				
DB	1081	GGTTCCCAAAAAGCTTC	ATTAAGGTCATATTTG	TTTAAAGATGCTAC	1140				
QY	1141	TTTCTTAAACTTTAAC	GATACAAATTAACCT	TAAGTGGGCACAGAG	1200				
DB	1141	TTTCTTAAACTTTAAC	GATACAAATTAACCT	TAAGTGGGCACAGAG	1200				
QY	1201	TATACACAGGACAGAT	GATGATTAATTAAC	CTTGAAGGTTGAAAG	1260				
DB	1201	TATACACAGGACAGAT	GATGATTAATTAAC	CTTGAAGGTTGAAAG	1260				
QY	1261	CTAGTTGAGAAAAAGG	CTCCCTTAGGTTACAA	TTTGTAGATACTCT	1320				
DB	1261	CTAGTTGAGAAAAAGG	CTCCCTTAGGTTACAA	TTTGTAGATACTCT	1320				
QY	1321	TTAGGAGATGGAGCC	CACTGATACAAATTA	TTTGTAGTTAAACCA	1380				
DB	1321	TTAGGAGATGGAGCC	CACTGATACAAATTA	TTTGTAGTTAAACCA	1380				
QY	1381	GAAAAATAAAGGTA	CTGAGTTGCTTTCA	ACAGGTGGTATTTG	1440				
DB	1381	GAAAAATAAAGGTA	CTGAGTTGCTTTCA	ACAGGTGGTATTTG	1440				
QY	1441	ATTATAGTGCATTTT	TAGTAATAGGACAGG	TATCGTGCTTTG	1500				
DB	1441	ATTATAGTGCATTTT	TAGTAATAGGACAGG	TATCGTGCTTTG	1500				
QY	1501	CGTTCCTTAA	1509						
DB	1501	CGTTCCTTAA	1509						

RESULT 2  
US-10-206-576-217  
; Sequence 217, Application US/10206576  
; Publication No. US20030017495A1  
; GENERAL INFORMATION:

APPLICANT: Choi et al.  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 497  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-R  
COMPUTER: Dell Latitude  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/206,576  
FILING DATE: 29-Jul-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/071,035  
FILING DATE: 1998-05-04  
APPLICATION NUMBER: US 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: US 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: US 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Hyman, Mark J.  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB369PID1  
INFORMATION FOR SEQ ID NO: 217:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1983 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 217:  
US-10-206-576-217  
Query Match 4.3%; Score 64.8; DB 13; Length 1983;  
Best Local Similarity 52.1%; Pred. No. 0.0012;  
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;  
QY 1180 ACAGAGCTTAATGCAACAGATATATACACAGGAGCAGATGGTATTAATTACCAATTACAGGC 1239  
DB 1660 ACAAAGCTGAAGCAACTACTTTTACACACACGCTGATGGATTAGTTGATATACAGGG 1719  
QY 1240 TTGAAGAAGGTACATATCTAGTTGAGAAAAAGGCTCCCTTAGGTTACAAATTTGTTA 1299  
DB 1720 CTTAAATACGGTACCTATTATTAGAAAGAACTGTAGCTCCTGATGATTATGCTTTGTTA 1779  
QY 1300 GATAACTCTCAGAGGTTATTTTAGGAGATGAGCGCACTGATACGACTAAATTCAGATAAC 1359  
DB 1780 ACAATCGGATGAATTTGTGTCATGCAATCATATGGCACAA-----CAGAAAC 1833  
QY 1360 CTTTGTAGTAAACCACTGTGAAATACAAAGGTACTGAGTTGCTTCAACAGGTGT 1419  
DB 1834 CTAGTTTACCAAGAAAAGTACCAACAAACAAAGGTACCTTACCTTCAACAGGTGGC 1893  
QY 1420 ATTGGTACAAATTTTCTACATTTAGTGCAATTTTAGTAAATAGGACAGGTATCGTG 1479  
DB 1894 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTATGACAGGAGTCTAC 1953  
QY 1480 CTTGTTCGTCGTCGTTTACGTTCTTAA 1509  
DB 1954 TTGCTAGACGTAGAAAAGAAATGCTTAA 1983

RESULT 3  
US-10-333-002-28  
; Sequence 28, Application US/10333002

Publication No. US20040071729A1  
GENERAL INFORMATION:  
APPLICANT: Aderson, Elisabeth  
APPLICANT: Bohnsack, John  
TITLE OF INVENTION: Group B Streptococcus Polypeptides Nucleic Acids and Therapeutic  
FILE OF INVENTION: Compositions and Vaccines Thereof  
FILE REFERENCE: 2511-1-001 (S7-0039)  
CURRENT APPLICATION NUMBER: US/10/333,002  
CURRENT FILING DATE: 2003-07-08  
PRIOR APPLICATION NUMBER: PCT/US01/24795  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 09/634,341  
PRIOR FILING DATE: 2000-08-08  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 28  
LENGTH: 2199  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
US-10-333-002-28  
Query Match 4.3%; Score 64.8; DB 12; Length 2199;  
Best Local Similarity 52.1%; Pred. No. 0.0013;  
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;  
QY 1180 ACAGAGCTTAATGCAACAGATATATACACAGGAGCAGATGGTATTAATTACCAATTACAGGC 1239  
DB 275 ACAAAGCTGAAGCAACTACTTTTACAAACACGCTGATGGATTAGTTGATATACAGGG 334  
QY 1240 TTGAAGAAGGTACATATCTAGTTGAGAAAAAGGCTCCCTTAGGTTACAAATTTGTTA 1299  
DB 335 CTTAAATACGGTACCTATTATTAGAAAGAACTGTAGCTCCTGATGATTATGCTTTGTTA 394  
QY 1300 GATAACTCTCAGAGGTTATTTTAGGAGATGAGCGCACTGATACGACTAAATTCAGATAAC 1359  
DB 395 ACAATCGGATGAATTTGTGTCATGCAATCATATGGCACAA-----CAGAAAC 448  
QY 1360 CTTTGTAGTAAACCACTGTGAAATACAAAGGTACTGAGTTGCTTCAACAGGTGT 1419  
DB 449 CTAGTTTCAACAGAAAAAGTACCAACAAACAAAGGTACTTACCTTCAACAGGTGGC 508  
QY 1420 ATTGGTACAAATTTTCTACATTTAGTGCATTTAGTAAATAGGACAGGTATCGTG 1479  
DB 509 AAAGGAATCTAGCTTTACTTAGGAAGTGGCGAGTCTTGCTACTTATTATGACAGGAGTCTAC 568  
QY 1480 CTTGTTCGTCGTCGTTTACGTTCTTAA 1509  
DB 569 TTGCTAGACGTAGAAAAGAAATGCTTAA 598  
RESULT 4  
US-09-070-927A-45  
; Sequence 45, Application US/09070927A  
; Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
Patrick J. Dillon  
Steven Barash  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:

```
/ APPLICATION NUMBER: US/09/070,927A
/ FILING DATE: 04-May-2000
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/   APPLICATION NUMBER: 60/046,655
/   FILING DATE: 1997-05-16
/   APPLICATION NUMBER: 60/044,031
/   FILING DATE: 1997-05-06
/   APPLICATION NUMBER: 60/066,009
/   FILING DATE: 1997-11-14
/ ATTORNEY/AGENT INFORMATION:
/   NAME: Kenley K. Hoover
/   REGISTRATION NUMBER: 40,302
/   REFERENCE/DOCKET NUMBER: PB369
/ TELECOMMUNICATION INFORMATION:
/   TELEPHONE: (301) 309-8504
/   TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 45:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 15614 base pairs
/     TYPE: nucleic acid
/     STRANDEDNESS: double
/     TOPOLOGY: linear
/ SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-070-927A-45

Query Match          4.3%; Score 64.8; DB 9; Length 15614;
Best Local Similarity 52.1%; Pred.No.0.0034;
Matches 172; Conservative 0; Mismatches 152; Indels 6; Gaps 1;

QY 1180 ACAGAGCTAATGCAACAGAAATATACAAACAGGAGCAGATGGTATATATACCAATTACAGGC 1239
Db 13724 ACAAAGCTGAAGCAACTACTTTTACAAACACGCTGATGGATTAGTTGATATACAGGC 13783

QY 1240 TTGAAGAAGGTACATCTACTCTAGTTGAGAAAAGGCTCCCTTAGTTACAATTGTTA 1299
Db 13784 CTTAAATACGGTACCTATTATTATTAGAGAAACTGTAGCTCCCTGATGATTATGCTTTGTTA 13843

QY 1300 GATAACTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 13844 ACAATCGGATTGAATTTGGTCAATGAACATCATATGGCACAA-----CAGAAAAC 13897

QY 1360 CTTTATAGTTAACCAACTGTGTAATAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 13898 CTAGTTTCACGAGAAAAGTACCAACAAACAAAGGTACCTTCAACAGGTGGC 13957

QY 1420 ATTGGTACAACTTTTCTACATTATAGTGCRAATTTTAGTAATAGGACAGGTATCGTG 1479
Db 13958 AAAGGAATCTACGTTTACTTTAGGAAGTGGCGCAGTCTTGCTACTTATTGAGGAGTCTAC 14017

QY 1480 CTTGTTGCTCGTCTGTTTACGTTCTTAA 1509
Db 14018 TTTGCTAGAGTAGAAAAGAAATGCTTAA 14047

RESULT 5
US-10-661-809-12
; Sequence 12, Application US/10661809
; Publication No. US2004010191A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR
; FILE REFERENCE: P07741US01/BAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1881
; TYPE: DNA
```

```
/ ORGANISM: Staphylococcus epidermidis
US-10-661-809-12

Query Match          4.2%; Score 63.2; DB 17; Length 1881;
Best Local Similarity 52.5%; Pred.No.0.0026;
Matches 165; Conservative 0; Mismatches 143; Indels 6; Gaps 1;

QY 1180 ACAGAAAGCTAATGCAACAGAAATATACAAACAGGAGCAGATGGTATATATACCAATTACAGGC 1239
Db 1561 ACAAAGCTGAAGCAACTACTTTTACAAACACGCTGATGGATTAGTTGATATACAGGC 1620

QY 1240 TTGAAGAAGGTACATCTACTCTAGTTGAGAAAAGGCTCCCTTAGTTACAATTGTTA 1299
Db 1621 CTTAAATACGGTACCTATTATTATTAGAGAAACTGTAGCTCCCTGATGATTATGCTTTGTTA 1680

QY 1300 GATAACTCTCAGAAAGGTTATTTTAGGAGATGGAGCCACTGATACGACTAATTCAGATAAC 1359
Db 1681 ACAATCGGATTGAATTTTGTGTCATATGAACATCATATGGCACAA-----CAGAAAAC 1734

QY 1360 CTTTATAGTTAACCCAACTGTTGAAAATAACAAAGGTACTGAGTTGCTTCAACAGGTGGT 1419
Db 1735 CTAGTTTCACGAGAAAAGTACCAACAAACAAAGGTACCTTCAACAGGTGGC 1794

QY 1420 ATTGGTACAACTTTTCTACATTATAGTGCAATTTTAGTAATAGGACAGGTATCGTG 1479
Db 1795 AAAGGAATCTACGTTTACTTTAGGAAGTGGCGCAGTCTTGCTACTTATTGAGGAGTCTAC 1854

QY 1480 CTTGTTGCTCGTCG 1493
Db 1855 TTTGCTAGACGTAG 1868

RESULT 6
US-10-282-122A-15988
; Sequence 15988, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
```





;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: 60/230,335  
;; PRIOR FILING DATE: 2000-09-06  
;; PRIOR APPLICATION NUMBER: 60/230,347  
;; PRIOR FILING DATE: 2000-09-09  
;; PRIOR APPLICATION NUMBER: 60/242,578  
;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/267,636  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 78614  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 40681  
;; LENGTH: 14067  
;; TYPE: DNA  
;; ORGANISM: Ureaplasma urealyticum  
US-10-282-122A-40681

Query Match 3.3%; Score 50.4; DB 13; Length 14067;  
Best Local Similarity 42.8%; Pred. No. 4;  
Matches 313; Conservative 0; Mismatches 416; Indels 3; Gaps 1;

QY 435 ATATGATATTTATTTCTAGCTCTTAATATGAGCTGTAATTTATGTTTACATC 494  
DB |||||  
QY 495 TGTAACTCCAAATGCTACTCATGAAAGAAATCTGATGCGACATGGGAGATGGTG 554  
DB |||||  
QY 8844 TCGAATTCACGAACCTTGAATTAATAAAAAATTCATTTGGTATATCTTCAATGACG 8903  
DB |||||  
QY 555 TGGAAAACTGTAGATCAAAAAAGCTACTCGGTTGGTGATACAGTCAATATACTATTAC 614  
DB |||||  
QY 8904 TGAAGAAATTTTAAAGTGATCAAAAAACATTTGCTTTGGGCTAATAACCACTATAGTTTCA 8963  
DB |||||  
QY 615 TTATAAGATGCGATCAATTTATCATGGTACAGAAAAAGTGTATCAATATGTTTATAAAGA 674  
DB |||||  
QY 8964 ATTGCTAATTTAAACATATTCGTAATATATACATTAAGAAAGTAAAGTAATTTAATGA 9023  
DB |||||  
QY 675 TACTATGCCATCTGCTCTGTAGTTGATTTGAACGAAGGCTCTTATGAAGTAACATTTAC 734  
DB |||||  
QY 9024 TGATAATAAAACATCAATTAATTTCCATTTAAAAAATGGAATAGCTGATTTGATTTTGT 9083  
DB |||||  
QY 735 TGATGGATCAGGAATATTACAACTCTAACTCAAGGTTCCGAAAAAGCAACTGGGAAGTA 794  
DB |||||  
QY 9084 TAATAAAACACAATCGATTAGTATAGTTTCAATAATTTGACCAATTTGCAAGAGCGAAAA 9143  
DB |||||  
QY 795 TAACTGTTAGAGGAAAAATAATTTTACGATTTACTATTTCGTTGGGAGCTACCAATAC 854  
DB |||||  
QY 9144 TAAATTACATCAACAATTTAGATTTATATTAATGATCTGTATATGTTTAAAGTA 9203  
DB |||||  
QY 855 TCCAAACCGGAATFACTCAAAATGGAGCTAATGATGACTTTTTTTATAGGGGAATAAATAC 914  
DB |||||  
QY 9204 TGAAGAAGAAGCAATCATTTAATTTTGGTGAGAAAT---TTATCTGTAAGCAAGCAAGTTAA 9260  
DB |||||  
QY 915 AATCAGATCACTTTATACAGGAGTATTAAGAGTGGAGCTAAACCGGTTACGCTGATTT 974  
DB |||||  
QY 9261 AATTGTAGGAGTCAAAAAATTTTAGAAAAACAACATTTCAATATCTAGCAATTTAAACCAAGA 9320  
DB |||||  
QY 975 ACCAGAAATACAAACATTCGACCAATCAACCCCAATCTAGCAATGATGACCCAGGTCA 1034  
DB |||||  
QY 9321 TACTATTATACAGTATTACTTTTATTAATAAACAATTAAGCTGCTACGAATTTGG 9380  
DB |||||  
QY 1035 AAAAGTAACAGTGAGGAGTGGTCAAAATTTACTATATAAAAAAATTTGATGGTTCCACAAAAAGC 1094  
DB |||||  
QY 9381 AATAGATAATAGCAATAAATAATATATACAACTACAAACATATATAATCCATTAAATTTGA 9440  
DB |||||  
QY 1095 TTCATTACAAGGTGCTATATTTGTTTAAAGAAATGCTACGGGTCAATTTCTAAACTTTAA 1154  
DB |||||

DB 9441 TAATAATTTTGTGTCATTTGGTCTCTGTTCAAGTGATACTAATAAACGTTATTACTTTGGA 9500  
QY 1155 CGATACAAATAA 1166  
DB |||||  
DB 9501 ATATAATAATAA 9512

RESULT 11  
US-10-087-464-42  
; Sequence 42, Application US/10087464  
; Publication No. US20030059436A1  
; GENERAL INFORMATION:  
; APPLICANT: Chishti, Athar  
; APPLICANT: Oh, Steven  
; APPLICANT: Liu, David  
; APPLICANT: Goel, Vikas  
; APPLICANT: Li, Xuerong  
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof  
; FILE REFERENCE: SI2377/019  
; CURRENT APPLICATION NUMBER: US/10/087,464  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 06/272,930  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 42  
; LENGTH: 3996  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
US-10-087-464-42

Query Match 3.3%; Score 50; DB 15; Length 3996;  
Best Local Similarity 47.5%; Pred. No. 2.6;  
Matches 149; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 366 TAATCTACACAGTTTCCACTGTTACTGAGTCAAAATTAATGATGGTACTGAGGTTATTAA 425  
DB |||||  
QY 426 TGTTCCTCCAAATGATGATTTATTTCTTAGCACTGTTAATATGAGCTGTTAATAT 485  
DB |||||  
QY 1941 TAATAATAATTTATGTTAACAATAATAACACAAATAATAAGGCAATAATAATAA 2000  
DB |||||  
QY 486 GGTTCATCTCTAATCTCCAAATGCTACTATTTCATGAAAGAAATCTGATGCGACATGGG 545  
DB |||||  
QY 2001 TGATGGAATGGTAGTAGTAATAATAATAATGATGATGATGACGAAGAAGAGGA 2060  
DB |||||  
QY 546 AGATGGTGGTGAAGAACTGTAGATCAAAAACGCTACTCCGGTTGGTGATACAGTCAAAATA 605  
DB |||||  
QY 2061 TGATGAAGATGATAACAATAATAATAATGATGATGATGATGATGATGATGATGATGATGAT 2120  
DB |||||  
QY 606 TACTATTACTTTATAGAAATGCGATCAATTTATCATGGTACAGAAAAAGTGTATCAATATGT 665  
DB |||||  
QY 2121 GGAAGTAATGATGAAGATAACGATGAGTATATAATAGTAATAGTATAGTATTAATAATGA 2180  
DB |||||  
QY 666 TATAAGGATACTA 679  
DB |||||  
DB 2181 AGAAAAAGATAGTA 2194

RESULT 12  
US-10-206-576-219  
; Sequence 219, Application US/10206576  
; Publication No. US20030017495A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et al.  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 497  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville

628 GTCAATTATCATGGTACAGAAA 650







```
; CURRENT APPLICATION NUMBER: US/10/373,667
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/181,585
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1037
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: comprising exons E, C, and A
US-10-373-667-3

Query Match          3.1%; Score 47.4; DB 16; Length 1037;
Best Local Similarity 47.2%; Pred. No. 4.8;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 218 TTCCTCAAGGTAAGAAGCTGAGTATATAAGCTTCAAGTAAATTTAAATCTCTTTTACGA 277
DB 611 TCCTTCATGTTAGAAAACCTGGCTTTACTACTACTACTACTACTACTACTACTACTACTA 670

QY 278 CAACTACTAATGGAGGAGAAACATATGTAACTAAAAAGATACCTGGTCAGCAAAATGAGA 337
DB 671 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730

QY 338 TTGGCAGATGGCTAAATCTATATACAGCTAATATGTTTCCCAATATGGATATTTATGTTTCTA 457
DB 731 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790

QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGGATATTTATGTTTCTA 457
DB 791 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850

QY 458 GCACGTGTTAATGAGGAGCTGAATATGTTTACATCTGTAACCTCAAATGCTACTATTC 517
DB 851 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910

QY 518 ATGAA 522
DB 911 CTGCA 915

RESULT 19
US-10-373-667-1
; Sequence 1, Application US/10373667
; Publication No. US20030235841A1
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 11000900101
; CURRENT APPLICATION NUMBER: US/10/373,667
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/181,585
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1159
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-373-667-1

Query Match          3.1%; Score 47.4; DB 16; Length 1159;
Best Local Similarity 47.2%; Pred. No. 5.1;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 218 TTCCTCAAGGTAAGAAGCTGAGTATATAAGCTTCAAGTAAATTTAAATCTCTTTTACGA 277
DB 423 TCCTTCATGTTAGAAAACCTGGCTTTACTACTACTACTACTACTACTACTACTACTACTA 482
```

```
QY 278 CAACTACTAATGGAGGAGAAACATATGTAACTAAAAAGATACCTGGTCAGCAAAATGAGA 337
DB 483 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542

QY 338 TTGGCAGATGGCTAAATCTATATACAGCTAATATGTTTCCCAATGTTCCACTGTTACTGAGT 397
DB 543 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602

QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGGATATTTATGTTTCTA 457
DB 603 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662

QY 458 GCACGTGTTAATGAGGAGCTGAATATGTTTACATCTGTAACCTCAAATGCTACTATTC 517
DB 663 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722

QY 518 ATGAA 522
DB 723 CTGCA 727

RESULT 20
US-10-373-667-2
; Sequence 2, Application US/10373667
; Publication No. US20030235841A1
; GENERAL INFORMATION:
; APPLICANT: Ranum, Laura P.W.
; TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 11000900101
; CURRENT APPLICATION NUMBER: US/10/373,667
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: US/09/181,585
; PRIOR FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1471
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cDNA
; OTHER INFORMATION: comprising exons D, C, B, and A
US-10-373-667-2

Query Match          3.1%; Score 47.4; DB 16; Length 1471;
Best Local Similarity 47.2%; Pred. No. 5.7;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 218 TTCCTCAAGGTAAGAAGCTGAGTATATAAGCTTCAAGTAAATTTAAATCTCTTTTACGA 277
DB 1045 TCCTTCATGTTAGAAAACCTGGCTTTACTACTACTACTACTACTACTACTACTACTACTA 1104

QY 278 CAACTACTAATGGAGGAGAAACATATGTAACTAAAAAGATACCTGGTCAGCAAAATGAGA 337
DB 1105 CTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1164

QY 338 TTGGCAGATGGCTAAATCTATATACAGCTAATATGTTTCCCAATGTTCCACTGTTACTGAGT 397
DB 1165 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1224

QY 398 CAAATAATGATGGTACTGAGGTATTAAATGTTTCCCAATATGGATATTTATGTTTCTA 457
DB 1225 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1284

QY 458 GCACGTGTTAATGAGGAGCTGAATATGTTTACATCTGTAACCTCAAATGCTACTATTC 517
DB 1285 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1344

QY 518 ATGAA 522
DB 1345 CTGCA 1349
```

RESULT 21

US-09-815-242-4843

Sequence 4843, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zvekind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4843

LENGTH: 4047

TYPE: DNA

ORGANISM: Staphylococcus aureus

US-09-815-242-4843

Query Match 3.1%; Score 47.4; DB 9; Length 4047;

Best Local Similarity 46.1%; Pred. No. 9.5;

Matches 159; Conservative 0; Mismatches 186; Indels 0; Gaps 0

QY 162 AGAATAGATATGCAATGTATCTGATTGCAATATAAGATGAGCTTCTTATTTTAATTC 221

DB 333 AGAAAAGAAAGTGTACATCTACCACCTGGAATTAAGTTGAAGTTTCAACTGCCAAATC 392

QY 222 TCAAGGTAAAGAGCTGTAGTATAAAGCTTCAAAGCTTTAACTCTCTTTTACGACAC 281

DB 393 AGATGAGCAAGCTTCACCAAAATCTACGAAATGAAGATTTAAACACATAACAACATATAAG 452

QY 282 TACTAATGAGGAGGAGCAACATATGTAACATAAAAGATCTCGGTGAGCAAAATGAGATTGC 341

DB 453 TAATCAAGAGCGTTACACCTGATTTGCAAGAGAAATAAATCAGTGGTAAATGTTCAACC 512

QY 342 GACATGGGCTAAATCTATATCAGCTAAATCTACACAGTTTCCACTGTGTACTGTAGTCAAA 401

DB 513 AACTAATGAGGAAAAACAAAAGGTAGATGCCAAAATCGAATCAACTACATTAATGTTAA 572

QY 402 TAATGATGCTAGTCTAGGTTTATTAAATGTTTCCCAATATGATATTAATGTTTCTAGCAC 461

DB 573 AAGTAGTCTATCAGAGTAATGATGAAGAACTCTTGTGTAAACATAGTAATTAATTA 632

QY 462 TGTTAATAATGAGCTGTAAATTAATGTTTACATCTGTAACTCCAAA 506

DB 633 TGAATAATACAGATATCATTTTGCAAAAAGTACAGACCTAA 677

RESULT 23  
US-08-781-986A-454/c  
; Sequence 454, Application US/08781986A  
; Publication No. US20030054436A1

RESULT 22  
US-09-815-242-9039  
: Sequence 9039, Application US/09815242

```
;
;
; GENERAL INFORMATION:
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 454:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4358 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-781-986A-454
;
; Query Match 3.1%; Score 47.4; DB 8; Length 4358;
; Best Local Similarity 46.1%; Pred. No. 9.8;
; Matches 159; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
;
; QY 162 AGAATAGATATGCAATGATCTGATTCGAATAAAGATGAGCTTCTTATTTAATTC 221
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 631 AGAAAAAGAAAGTGATCAATCTACCACCTGGAATAAAGTTGAAGTTTCAACTGCCAAATC 572
;
; QY 222 TCAAGGTTAAGAAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTACGACAC 281
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 571 AGATGAGCAAGCTTACCACCAATCTACGAAATGAAAGTTTAAACACTAAACAACTATTAAG 512
;
; QY 282 TACTAATGAGGAGGAGAACATATGTAATCTAAAAAGATGACTCGCTCAGCAAAATGAGATTGC 341
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 511 TAATCAAGACGTTTACACCTGATTTGCAAGAGANAATAATCAGTGGTAATGTTCAACC 452
;
; QY 342 GACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACTGTTTCTAGTCAAA 401
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 451 AACTAATGAGGAAAAAAGAGTAGATGCCAAAATCTGAATCAACTACATTAAATGTTAA 392
;
; QY 402 TAATGATGCTAGCTGAGGTTATTAATGTTTCCCAATATGATATTTATGTTTCTAGC 461
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 391 AAGTATGCTATCAAGAGTAATGATGAAACTCTTTGTTGATAACAAATAGTAATCAATAA 332
;
; QY 462 TGTAAATGAGCTGTAATTTATGTTTACACTCTGTAACCTCAAA 506
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 331 TGAATAATGAGATATCATTTTGGCCAAAAGTACAGCACTAA 287
;
; RESULT 24
; US-10-329-624-454/c
; Sequence 454, Application US/10329624
; Publication No. US2004004307A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
;
;
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
;
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/329,624
; FILING DATE: 27-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/956,171
; FILING DATE: October 20, 1997
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 454:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4358 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-10-329-624-454
;
; Query Match 3.1%; Score 47.4; DB 13; Length 4358;
; Best Local Similarity 46.1%; Pred. No. 9.8;
; Matches 159; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
;
; QY 162 AGAATAGATATGCAATGATCTGATTCGAATAAAGATGAGCTTCTTATTTAATTC 221
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 631 AGAAAAAGAAAGTGATCAATCTACCACCTGGAATAAAGTTGAAGTTTCAACTGCCAAATC 572
;
; QY 222 TCAAGGTTAAGAAAGCTGAGTATAAAGCTTCAACTGATTTTAAATCTCTTTTACGACAC 281
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 571 AGATGAGCAAGCTTACCACCAATCTACGAAATGAAAGTTTAAACACTAAACAACTATTAAG 512
;
; QY 282 TACTAATGAGGAGGAGAACATATGTAATCTAAAAAGATGACTCGCTCAGCAAAATGAGATTGC 341
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 511 TAATCAAGACGTTTACACCTGATTTGCAAGAGANAATAATCAGTGGTAATGTTCAACC 452
;
; QY 342 GACATGGGCTAAATCTATATCAGCTAATACTACACAGTTTCCACTGTTTCTAGTCAAA 401
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 451 AACTAATGAGGAAAAAAGAGTAGATGCCAAAATCTGAATCAACTACATTAAATGTTAA 392
;
; QY 402 TAATGATGCTAGCTGAGGTTATTAATGTTTCCCAATATGATATTTATGTTTCTAGC 461
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 391 AAGTATGCTATCAAGAGTAATGATGAAACTCTTTGTTGATAACAAATAGTAATCAATAA 332
;
; QY 462 TGTAAATGAGCTGTAATTTATGTTTACACTCTGTAACCTCAAA 506
; DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; 331 TGAATAATGAGATATCATTTTGGCCAAAAGTACAGCACTAA 287
;
; US-10-329-624-454/c
; Sequence 454, Application US/10329624
; Publication No. US2004004307A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
```

Db 1949 CAAATTATGATAATCAAAATTATGATAATCAAAATTATAAAT 1992

RESULT 26

US-10-282-122A-34844

Sequence 34844, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 34844

LENGTH: 2801

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-10-282-122A-34844

Query Match 3.1%; Score 46.8; DB 13; Length 2801;

Best Local Similarity 47.6%; Pred. No. 11;

Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps

QY 289 GGAGGGAGAACATATGTAACATAAAAAAGATACTGCGTCAGCAAAATGAGATTGCGACATGTC

DB 1111 GAAGGTGGACAACTACACAGGTTCAAAAGTATATCCATGGATTGCTGAGACTAGAGGTTT

QY 349 GCTAAATCTATATCAGCTAATATCTACACCAAGTTTCCACTGTTACTGAGTCAAAATATGTA

DB 1171 ACACGAGTATAGGATTAATTTCCACCACCACCACCATGATATTTATTCCAATTGAGGA

QY 409 GGTACTGAGGTTTATATGTTTCCCAATATGGAATATTTATTGTTTCTAGCACTGTAA

DB 1231 TTGACAGAGCTCATTCATGATTTTAAAAAATGCAAAATAGAGAGCTGATATTTCAGTTAA

QY 469 AATGGAGCTGTAATATTATGGTTTACATCTGTAACCTCCAAATGCTACTATTCATGAAAGAA

DB 1281 CTTGTATCAAAACTGCGTTTGGAACTATAGCTTCAGGGTAGCTAAAGCTTTCGCCGCA



```
QY 402 TAATGATGGTACTGAGGTTAATTAATGTTCCCAATATGGATATTAATTAATGTTTCTAGCAC 461
Db |||||
573 AAGTGATGCTATCAAGAGTAATGCTGAACACTCTGTTGATAACAATAGTAATTCAAATAA 632
QY 462 TGTTAATAATGGAGCTGTTAATTAATGTTTACATCTGTAATCTGTAATCTCAATGCTACTATTTCATGA 521
Db |||||
633 TGAATAATAATGAGATATCATTTTGGCAAAAAGTACAGCACCTTAAAGTTTGAATACAAG 692
QY 522 AAAGAATACTGATGCGACA 540
Db |||||
693 AATGCGTATGGCACAATA 711

RESULT 29
US-10-398-221-9/c
; Sequence 9, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 9
; LENGTH: 684707
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

Query Match 3.0%; Score 45.8; DB 16; Length 684707;
Best Local Similarity 44.9%; Pred. No. 2.6e+02;
Matches 173; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 956 AACCAGGTTCCAGCTGATTTACCAAGAAAATACAAACATTCGACCATCAACCCCAATACTA 1015
Db |||||
175405 AAGAGATTATCAATATAAAACGAAGAAGGAAATTTAAAAATGTTTAAAGTAGATCATA 175346
QY 1016 GCAATGATGACCCAGGTCAAAAGTAACAGTGAGGGATGGTCAAAATTACTATAAAAAAA 1075
Db |||||
175345 ATGATGTTTTCAAAATGGAGTAGAAAATGGTACGTATGAGGTGGTTTTATACACGCAA 175286
QY 1076 TTGATGTTCCACAAAGCTTCATTACAGGTGCTATATTTGTTTAAAGATGCTACGG 1135
Db |||||
175285 ATGAAGATGGCAAAAAACGGAGCGGAGTTCAATTAATTTGAATTTAATTTTCGTAATG 175226
QY 1136 GTCAATTTCTAAACTTTAACGATACAAATTAACGTTGGGCGACAGAAAGCTAATGCAA 1195
Db |||||
175225 ATGTAATCAAAATTCGAATCGCATATTTTCCAGGATATGGAAGCAAAAGCAA 175166
QY 1196 CAGAATATACACAGGAGCAGATGGTATAATTAACATTAACAGGCTTGAAGAAGGTACAT 1255
Db |||||
175165 CAAATGAATATAGTCAAAACGGCATTTAAATACAAATCGCTAAAGCAATCCAATTTACCTAACG 175106
QY 1256 ACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAACTCTCAGAAG 1315
Db |||||
175105 GCAAGATTAATAATCATTTGGATGAATTTTAAAGACCTGTTAACTAAGACATGCCAAG 175046
QY 1316 TTATTTTAGGAGATGGAGCCACTGA 1340
Db |||||
175045 TTACTGTGAAAATGAAGACTGA 175021
```

```
RESULT 30
US-10-398-221-2058/c
; Sequence 2058, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2058
; LENGTH: 3011208
; TYPE: DNA
; ORGANISM: Listeria innocua
; US-10-398-221-2058

Query Match 3.0%; Score 45.8; DB 16; Length 3011208;
Best Local Similarity 44.9%; Pred. No. 4.6e+02;
Matches 173; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 956 RACCAGGTTCCAGTCGATTTACCAGAAAATACAAACATTCGACCATCAACCCCAATACTA 1015
Db |||||
2612988 AAGAGATTATCAATATAAAACGAAGAAGAAATTTAAAAATGTTTAAAGTAGATCATA 2612929
QY 1016 GCAATGATGACCCAGGTCAAAAGTAACAGTGAGGGATGGTCAAAATTACTATAAAAAAA 1075
Db |||||
2612928 ATGATGTTTTCACAAATGGAGTAGAAAATGGTACGTATGAGGTGGTTTTATACAACGCA 2612869
QY 1076 TTGATGGTTCCACAAAAGCTTCATTACAAGGTGCTATATTTGTTTAAAGATGCTACGG 1135
Db |||||
2612868 ATGAAGATCGCAAAAAACGGAGCGGAGTTTCATTAAATATTGAATTTAATTTTCGTAATG 2612809
QY 1136 GTCAATTTCTAAACTTTAACGATACAAATTAACGTTGGGCGACAGAAAGCTAATGCAA 1195
Db |||||
2612808 ATGTAATCAAAATTCAGATGCGCATATTTTCCAGGATATGGAAGCAAAAGCAA 2612749
QY 1196 CAGAATATACACAGGAGCAGATGGTATAATTAACATTACAGGCTTTGAAAGAAGGTACAT 1255
Db |||||
2612748 CAAATGAATATATAGTCAAAACGGCATTTAAATACAAATCGCTAAAGCAATCCAATTTACCTAACG 2612689
QY 1256 ACTATCTAGTTGAGAAAAGGCTCCCTTAGGTTACAAATTTGTTAGATAACTCTCAGAAG 1315
Db |||||
2612688 GCAAGATTTATAATACATTTGGATGAATTTTAAAGACCTGTTAACTAAGACATGCCAAG 2612629
QY 1316 TTATTTTAGGAGATGGAGCCACTGA 1340
Db |||||
2612628 TTACTGTGAAAATGAAGACTGA 2612604

Search completed: July 27, 2004, 03:05:42
Job time : 739 secs
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 09:58:15 ; Search time 54 Seconds  
(without alignments)  
2626.645 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKMIQSLVSLVAFGMV.....GALLVIGIVIVARRLRIS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : A\_Geneseq\_29Jan04.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2578	100.0	502	4	AAY72357 Virulent
2	2578	100.0	502	5	AAY77626 S. agalac
3	339	15.5	525	5	ABP65654 Bifidobac
4	334	13.0	627	2	AAY00118 Enterococ
5	334	13.0	627	5	ABP43337 E faecali
6	334	13.0	627	6	ABU88365 E. faecal
7	334	13.0	627	6	ABU13616 Enterococ
8	318	12.3	664	7	ADC97625 E. faeciu
9	279.5	10.8	705	5	ABP27284 Streptoco
10	274.5	10.6	560	2	AAY00119 Enterococ
11	274.5	10.6	560	5	ABP43338 E faecali
12	274.5	10.6	560	6	ABU13617 Enterococ
13	274.5	10.6	560	6	ABU88366 E. faecal
14	265.5	10.3	665	3	AAY81627 Streptoco
15	265.5	10.3	666	3	AAY81627 Streptoco
16	258.5	10.0	614	5	ABP54071 Lactococc
17	244.5	9.5	554	5	ABP29802 Streptoco
18	244.5	9.5	581	5	ABP27373 Streptoco
19	234.5	9.1	341	7	ADC95757 E. faeciu
20	232	9.0	793	5	ABP47318 Listeria
21	232	9.0	793	6	ABU132765 Protein e
22	222.5	8.6	688	4	AAM16315 Peptide #
23	222.5	8.6	688	4	ABP35307 Peptide #
24	222.5	8.6	688	4	AAM28810 Peptide #
25	222.5	8.6	688	4	ABB30137 Peptide #

RESULT 1

AAY72357  
ID AAY72357 standard; protein; 502 AA.

#### ALIGNMENTS

26	222.5	8.6	688	4	ABB20749	Abb20749	Protein #
27	222.5	8.6	688	4	AAM68511	Aam68511	Human bon
28	222.5	8.6	688	4	AAM56138	Aam56138	Human bra
29	222.5	8.6	688	4	ABG50180	Abg50180	Human liv
30	222.5	8.6	688	4	AAM04053	Aam04053	Peptide #
31	222.5	8.6	688	5	ABG38092	Abg38092	Human pep
32	220	8.5	560	5	AAY77628	Aay77628	S. agalac
33	220	8.5	563	4	AAY72358	Aay72358	Virulent
34	219	8.5	1222	3	AAAB01830	Aab01830	H. influe
35	219	8.5	1228	3	AAAB01828	Aab01828	Haemophil
36	217	8.5	1612	5	ABBA47333	Abb47333	Listeria
37	217	8.5	562	5	ABBA47328	Abb47328	Listeria
38	212	8.2	645	6	ABU45710	Abu45710	Protein e
39	212	8.2	2551	6	ABU46159	Abu46159	Protein e
40	208.5	8.1	660	7	ADC97088	Adc97088	E. faeciu
41	205.5	8.0	1027	7	ADC97048	Adc97048	E. faeciu
42	201.5	7.8	887	3	AAY81626	Aay81626	Streptoco
43	200.5	7.8	893	6	ABU00820	Abu00820	S. pneumo
44	200	7.8	1440	5	ABBS4801	Abb54801	Lactococc
45	199	7.7	508	7	ADC95704	Adc95704	E. faeciu
46	199	7.7	1530	5	ABBA47329	Abb47329	Listeria
47	198.5	7.7	724	6	ABU17871	Abu17871	Protein e
48	198	7.7	596	3	AAY99408	Aay99408	Human PRO
49	198	7.7	596	4	ABG66157	Abg66157	Protein o
50	198	7.7	596	4	AAU29178	Aau29178	Human PRO
51	198	7.7	596	4	ABAB7575	Abab7575	Human PRO
52	198	7.7	596	5	ABG95900	Abg95900	Human sec
53	198	7.7	596	6	ABU58554	Abu58554	Human PRO
54	198	7.7	596	6	ABU88102	Abu88102	Novel hum
55	198	7.7	596	6	ABU84417	Abu84417	Human sec
56	198	7.7	596	6	ABR66291	Abri66291	Human sec
57	198	7.7	596	6	ABR65681	Abri65681	Human sec
58	198	7.7	596	6	ABU99621	Abu99621	Human sec
59	198	7.7	596	6	ABU82860	Abu82860	Human PRO
60	198	7.7	596	6	ABU89981	Abu89981	Novel hum
61	198	7.7	596	6	ABR68230	Abri68230	Human sec
62	198	7.7	596	6	ABU96283	Abu96283	Novel hum
63	198	7.7	596	6	ABU92714	Abu92714	Human sec
64	198	7.7	596	6	ABO08791	Ab008791	Human sec
65	198	7.7	596	6	ABO02843	Ab002843	Human sec
66	198	7.7	596	6	ABR74997	Abri74997	Human sec
67	198	7.7	596	6	ABR94759	Abri94759	Human sec
68	198	7.7	596	6	ABU85732	Abu85732	Human PRO
69	198	7.7	596	6	ABU98892	Abu98892	Novel hum
70	198	7.7	596	6	ABU98107	Abu98107	Novel hum
71	198	7.7	596	6	ABU91813	Abu91813	Novel hum
72	198	7.7	596	6	ABU89506	Abu89506	Human PRO
73	198	7.7	596	6	ABU86347	Abu86347	Human sec
74	198	7.7	596	6	ABU67560	Abu67560	Human sec
75	198	7.7	596	6	ABU80588	Abu80588	Human PRO
76	198	7.7	596	6	ABU90925	Abu90925	Novel hum
77	198	7.7	596	6	ABO33984	Ab033984	Human sec
78	198	7.7	596	6	ABR99506	Abri99506	Human sec
79	198	7.7	596	6	ABR98896	Abri98896	Human sec
80	198	7.7	596	6	ABO16419	Ab016419	Human sec
81	198	7.7	596	6	ABR92319	Abri92319	Human sec
82	198	7.7	596	6	ABR18960	Abri18960	Human sec
83	198	7.7	596	6	ABR78381	Abri78381	Human sec
84	198	7.7	596	6	ABU72001	Abu72001	Novel hum
85	198	7.7	596	6	ABU85117	Abu85117	Novel hum
86	198	7.7	596	6	ABO00256	Ab000256	Novel hum
87	198	7.7	596	6	ABO11588	Ab011588	Human sec
88	198	7.7	596	6	ABO02233	Ab002233	Human sec
89	198	7.7	596	6	ABU88807	Abu88807	Novel hum
90	198	7.7	596	6	ABU83502	Abu83502	Human sec

```

XX AC AAY72357;
XX DT 24-APR-2001 (first entry)
XX DE Virulent group B Streptococcus agalactiae spbl protein.
XX KW Type III virulent group B; spbl; cell wall bound protein; antibacterial;
XX KW immunisation; group B streptococci; GBS infection; pneumonia; meningitis;
XX KW endocarditis; Osteoarticular infection; bacteraemia; vaccine.
XX OS Streptococcus agalactiae.
XX FH Key Location/Qualifiers
XX FT Region 1..6
XX FT /note= "This region is hydrophilic and has six basic
XX FT amino acid residues"
XX FT Peptide 7..29
XX FT /label= Signal peptide
XX FT /note= "This region is hydrophobic and is rich in
XX FT proline"
XX FT Protein 30..502
XX FT /label= mature_spbl_protein
XX PN MO200078787-A1.
XX XX
XX PD 28-DEC-2000.
XX XX
XX PF 21-JUN-2000; 2000WO-US017082.
XX XX
XX PR 21-JUN-1999; 99US-0140084P.
XX XX
XX PA (UTAH ) UNIV UTAH RES FOUND.
XX XX
XX PI Adderson E, Bohnsack J;
XX XX
XX DR WPI; 2001-102693/11.
XX DR N-PSDB; AAD02390.
XX XX
XX PT Polynucleotide from spbl and 2 genes derived from virulent Group B
XX PT streptococci, polypeptide encoded by the polynucleotide useful as vaccine
XX PT for immunizing a mammal against the streptococcal infection.
XX XX
XX PS Claim 56; Page 23-24; 34pp; English.
XX XX
XX CC The present sequence is spbl protein from type III virulent group B
XX CC Streptococcus agalactiae. The spbl protein has the characteristics of a
XX CC cell wall bound protein and has antibacterial activity. The N-terminus
XX CC of the spbl protein is a hydrophilic, basic stretch of 6 amino acids
XX CC followed by a 23 amino acid hydrophobic, proline rich core, consistent
XX CC with a signal peptide. The hydrophilic mature protein terminates in
XX CC atypical LPxTG domain that immediately precedes a hydro- phobic 20 amino
XX CC acid core and a short, basic hydrophilic terminus. The spbl protein are
XX CC used as a vaccine to immunise mammals against group B Streptococci (GBS)
XX CC infection (e.g. bacteraemia, pneumonia, meningitis, endocarditis and
XX CC osteoarticular infections). Determination of the gene products specific
XX CC to type III-3 GBS is useful for diagnosing mammals infected or colonised
XX CC by virulent GBS
XX XX
XX SQ Sequence 502 AA;
XX XX
Query Match 100.0%; Score 2578; DB 4; Length 502;
Best Local Similarity 100.0%; Pred. No. 1e-146;
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 MKKKMQLSLVASLAFGMVSPVTPAFAPAEAGTITVQDTOKGATYKAVKVPDAEDNAN 60
XX DB 1 MKKKMQLSLVASLAFGMVSPVTPAFAPAEAGTITVQDTOKGATYKAVKVPDAEDNAN 60
XX QY 61 VSDSNKDGASLYLPQGEAEYKASTDFNSLFTTTTNGGRYTVTKDTSANAEIATWAKSI 120
XX DB 61 VSDSNKDGASLYLPQGEAEYKASTDFNSLFTTTTNGGRYTVTKDTSANAEIATWAKSI 120

```

```

QY 121 SANTTPSVSTVESNNDGTEVINVSQYGYVYSSVTNNGAVIMVTSVTPNATHEKNTDAT 180
DB |||||
DB 121 SANTTPSVSTVESNNDGTEVINVSQYGYVYSSVTNNGAVIMVTSVTPNATHEKNTDAT 180
QY 181 WGDGGGKTVDQKTSYSGDVTVKYTIYKNAVNYHGTEKYQYVYIKDTMPSASVVDLNEGSY 240
DB |||||
DB 181 WGDGGGKTVDQKTSYSGDVTVKYTIYKNAVNYHGTEKYQYVYIKDTMPSASVVDLNEGSY 240
QY 241 EVTIIDGSGNIITTLTQGSSEKATGKYNLLEENNFTITIPWAATNPTGNTQNGANDDPFY 300
DB |||||
DB 241 EVTIIDGSGNIITTLTQGSSEKATGKYNLLEENNFTITIPWAATNPTGNTQNGANDDPFY 300
QY 301 KGINTITVTYTGVLXSGAKPGSADLPENTNIATINPNTSNDPQKQVTVRQGIITIKKID 360
DB |||||
DB 301 KGINTITVTYTGVLXSGAKPGSADLPENTNIATINPNTSNDPQKQVTVRQGIITIKKID 360
QY 361 GSTKASLOGAIFVLKNATGQFLNFNTNNVNGTEANATEYTTGADGIIITITGLKEGTYY 420
DB |||||
DB 361 GSTKASLOGAIFVLKNATGQFLNFNTNNVNGTEANATEYTTGADGIIITITGLKEGTYY 420
QY 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPSTGGIGTTIFY 480
DB |||||
DB 421 LVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPSTGGIGTTIFY 480
QY 481 IIGAILVITGAGIVLVARRRLRS 502
DB |||||
DB 481 IIGAILVITGAGIVLVARRRLRS 502

RESULT 2
AAU77626
ID AAU77626 standard; protein; 502 AA.
XX AC AAU77626;
XX DT 05-JUN-2002 (first entry)
XX DE S. agalactiae SbpI protein.
KW Extracellular matrix adhesion; Ema; group B streptococcus; GBS; vaccine;
KW SbpI; Sbp2; Rib; Lmb; C5a-ase; C protein alpha antigen;
KW neonatal bacterial infection.
XX OS Streptococcus agalactiae.
XX PN WO200212294-A2.
XX PD 14-FEB-2002.
XX PF 08-AUG-2001; 2001WO-US024795.
XX PR 08-AUG-2000; 2000US-00634341.
XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX PA (UTAH ) UNIV UTAH RES FOUND.
XX PI Adderson E, Bohnsack J;
XX DR WPI; 2002-257465/30.
XX DR N-PSDB; ABK11585.
XX PT New streptococcal matrix adhesion (Ema) polypeptides, useful as vaccines,
XX PT particularly for treating or preventing infections by virulent forms of
XX PT streptococci.
XX PS Example 1; Page 150-152; 177pp; English.
XX CC The invention relates to isolated streptococcal polypeptides, which
XX CC comprise the Streptococcus agalactiae extracellular matrix adhesion (Ema)
XX CC polypeptides EmaA, EmaB, EmaC, EmaD or EmaE, and their encoding nucleic
XX CC acids. Also included are the a vaccine comprising an Ema protein, an anti
XX CC -Ema antibody, a pharmaceutical composition comprising one or more Ema
XX CC proteins plus an anti Ema antibody, optionally in combination with at

```

CC least one antibody to a protein selected from Spb1 and Spb2, Rib, Lmb,  
CC Csa-ase or C protein alpha antigen, an immortal cell line producing an  
CC anti-Ema monoclonal antibody, a vector comprising the Ema nucleic acid, a  
CC host cell transformed with the vector, a nucleic acid vaccine comprising  
CC the vector and Ema homologous proteins and their encoding nucleic acids  
CC from additional bacterial species (*S. pneumoniae*, *S. pyogenes*,  
CC *Enterococcus faecalis* and *Corynebacterium diphtheriae*). The streptococcal  
CC polypeptides are useful as vaccines, particularly for treating or  
CC preventing infections by virulent forms of streptococci, especially group  
CC B streptococci (GBS) the most common cause of serious bacterial disease  
CC in neonates. The present sequence is the *S. agalactiae* Spb1 protein  
XX  
SQ Sequence 502 AA;

Query Match 100.0%; Score 2578; DB 5; Length 502;  
Best Local Similarity 100.0%; Pred. No. 1e-146;  
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKKKMIQSLVSLVAFGMVSPVPTIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNAN 60  
Db 1 MKKKMIQSLVSLVAFGMVSPVPTIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNAN 60  
Qy 61 VSDSNKDGASYLIPQKGAAYKASTDFNSLFTTTNGGRTYVTKKDTASANEIATWAKSI 120  
Db 61 VSDSNKDGASYLIPQKGAAYKASTDFNSLFTTTNGGRTYVTKKDTASANEIATWAKSI 120  
Qy 121 SANTPTPVSTVTSNNDGTEVINVSQYGYVSVTVNNGAVIMVTSVTPNATIHKNTDAT 180  
Db 121 SANTPTPVSTVTSNNDGTEVINVSQYGYVSVTVNNGAVIMVTSVTPNATIHKNTDAT 180  
Qy 181 WGDGGGKTVQKTSYSGVTYKTYTYKNAVNYHGTEKYQYVVKDTMPSASVDLNEGSI 240  
Db 181 WGDGGGKTVQKTSYSGVTYKTYTYKNAVNYHGTEKYQYVVKDTMPSASVDLNEGSI 240  
Qy 241 EVTTIDGSGNITTLTQSGEKATGKYNLEENNNFTITPWAATNPTGNTONGANDDFY 300  
Db 241 EVTTIDGSGNITTLTQSGEKATGKYNLEENNNFTITPWAATNPTGNTONGANDDFY 300  
Qy 301 KGINTITVYTVGLKSGAKPGSADLPENTNIATINPNTSDPQKQVVRDQGITIKKID 360  
Db 301 KGINTITVYTVGLKSGAKPGSADLPENTNIATINPNTSDPQKQVVRDQGITIKKID 360  
Qy 361 GSTKASLOGAIFVLKNATQFLNFDNTNNVENGTEANATEYTTGADGIIITITGLKEGTY 420  
Db 361 GSTKASLOGAIFVLKNATQFLNFDNTNNVENGTEANATEYTTGADGIIITITGLKEGTY 420  
Qy 421 LVEKKAPGLYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPSTGGIGTITFY 480  
Db 421 LVEKKAPGLYNLLDNSQKVLGDGATDTTNSDNLVNPVTENNKGTELPSTGGIGTITFY 480  
Qy 481 IIGAILVIGAGIVLVARRLRS 502  
Db 481 IIGAILVIGAGIVLVARRLRS 502

RESULT 3  
ID ABP65654 standard; protein; 525 AA.  
XX  
AC ABP65654;  
XX  
DT 19-NOV-2002 (first entry)  
XX  
DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:398.  
XX  
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KW antidiarrheic; antibacterial; inhibitor of salmonella; detection;  
KW identification; lactic acid bacteria; diarrhoea; pathogenic bacteria;  
KW rotavirus; food composition; pharmaceutical composition.  
XX  
OS Bifidobacterium longum.  
XX  
PN EP1227152-A1.

XX 31-JUL-2002.  
PD  
XX 30-JAN-2001; 2001EP-00102050.  
PF  
XX 30-JAN-2001; 2001EP-00102050.  
PR  
XX (NEST ) SOC PROD NESTLE SA.  
PA  
XX WPI; 2002-668397/72.  
DR  
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful as  
PT a probe or primer for detecting and/or identifying Bifidobacterium longum  
PT in a biological sample.  
XX  
PS Claim 3; SEQ ID NO 398; 80pp; English.  
XX

CC The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
CC least 90% identity or which hybridises with the sequences given in  
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a  
CC fusion protein, comprising a sequence selected from 1097 sequences given  
CC in ABQ65258 to ABQ65354 ligated in frame to a polynucleotide encoding a  
CC heterologous polypeptide. (I) has antidiarrheic and antibacterial  
CC activities, and can be used as an inhibitor of *Salmonella*. (I) (which is  
CC a probe) is useful for the detection and/or identification of  
CC Bifidobacterium longum in a biological sample. A carrier containing the  
CC lactic acid bacterium Bifidobacterium longum NCC2705 (NCM 1-2618) can be  
CC used for preventing and/or treating diarrhoea brought about by pathogenic  
CC bacteria and/or rotavirus. The carrier is a food composition selected  
CC from milk, yogurt, curd, cheese, fermented milks, milk based fermented  
CC products, ice-creams, fermented cereal based products, milk based  
CC powders, infant formula, pet food or a pharmaceutical composition  
CC selected from tablets, liquid bacterial suspensions, dried oral  
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
CC (I) is useful in DNA arrays or chips to carry out analysis of the  
CC expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent  
CC Bifidobacterium related nucleotide sequences given in the Sequence  
CC Listing from the present invention but not mentioned further within the  
CC specification. N.B. The sequence data for this patent is not represented  
CC in the printed specification but is based on sequence information  
CC supplied by the European Patent Office

Sequence 525 AA;

Query Match 15.5%; Score 399; DB 5; Length 525;  
Best Local Similarity 27.3%; Pred. No. 7.3e-16;  
Matches 160; Conservative 72; Mismatches 198; Indels 156; Gaps 26;

Qy 3 KKMIIQSLVSLVAFGMVSPVPTIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNA 59  
Db 6 RKVAAGVLAAATMLGIAGLGTASAEATGTLVTSSDAFNGKKNVAYQMFSPDAA 65  
Qy 60 NVSDSNKDGASYLIPQKGAAYKASTDFNSLFTTTNGGRTYVTKKDTAS----- 109  
Db 66 -----GQATYTLASPDWNPFTNADLGITGTVTDANVSEKAYDYVYSL 109  
Qy 110 -----ANEIATWAKSISANTPTPVSTVTSNNDGTEVINVSQY-----GYYY 150  
Db 110 NQDMKLAEFATKASNWAKKAGNNVSAVNPAAATKASG-----SQYVATFANILPLGYIV 163  
Qy 151 VS-----STVN-----NGAVIMVTSVTPNATIH-----EKNTDA-TWGDG---GGKTVDQK 192  
Db 164 VSPAGSTSNKRHTDAMLVNVTATTKNLKSEYPTVDKTDADKGDGSAIGSKVNFQL 223  
Qy 193 TYSVGDVTYKTYTYKNAVNYHGTEKYQYVVKDTMPSASVDLNEGSEVITITGSGNIT 252  
Db 224 KSKVPDTSVT-----NYVFKIVDTLSAG--LDFFNN---DVTVKVGDATLT 264  
Qy 253 TLTOGSEKATGKYNLEENNNFTITIPWAATNPTGNTONGANDDFYKGINITITVYTG 312  
Db 265 ATTDYSVTTRGK-----TVTI-----DLSNVYKTDNASKAGKGLIVTYS 304

```
QY 313 VLKSGAKPGSADL--PENTNIATI--NPNTSNDPQGVTVRQGIT---IKKI--DGS 362
Db 305 TLNENAFVGTTPDQNNPGLNSAKVQVSNPGSEBENIGESTPSETHSYTFNFKIKIYKEGD 364
QY 363 TKASLOGAIFVLKNATGQFLNF-----NDTNNVWGTEANATEYTTGADGII 409
Db 365 TENALAGAKFQLLSDSKTVISLVKSDNVYRPAKTSDDDEV-----TEVETATGII 416
QY 410 TITGLKEGYIYLVKKAIPGLYNLLDNSQKVIILGDGATDTNS--DNLLN---PT----- 459
Db 417 EFTGLKAGTYLKETFPAGKYNKLSDPVKVTINATINKTGALESWTVNGSAPTADVTVP 476
QY 460 ---VENNKGTETLPGSTGGITTFYIIGAILV-IGAGIVLVARRR 501
Db 477 VKIENKKGALLPDTGGMGTFLTFVFGVILVALGAGWYVKSNRKS 522

RESULT 4
AAI00118
ID AAY00118 standard; protein; 627 AA.
XX
AC AAY00118;
DT
DT
XX
XX Enterococcus faecalis protein EF058.
DE
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW detection; attenuation; antigenic.
XX
XX Enterococcus faecalis.
OS
XX WO9805054-A2.
XX
XX 12-NOV-1998.
XX
XX 04-MAY-1998; 98WO-US008959.
XX
XX 06-MAY-1997; 97US-0044031P.
XX
XX 16-MAY-1997; 97US-0046655P.
XX
XX 14-NOV-1997; 97US-0066009P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Bailey C, Hromockyj A;
XX
XX WPI; 1999-070095/06.
XX
XX N-PSDB; AAX20108.
XX
XX New isolated Enterococcus faecalis polynucleotides - used to develop
XX products for the detection of Enterococcus and for use in vaccines for
XX prevention or attenuation of Enterococcus infection.
XX
XX Claim 9; Page 133; 301pp; English.
XX
XX The present sequence represents a protein isolated from Enterococcus
XX faecalis. The present invention describes genes, proteins and antigenic
XX polypeptides isolated from E. faecalis. The proteins can be used in
XX vaccines for preventing or attenuating an infection caused by a member of
XX the Enterococcus genus in an animal. They can also be used for detecting
XX Enterococcus antibodies in a sample. The nucleotide sequences can be used
XX for detecting Enterococcus nucleic acids. Products from the present
XX invention can also be used for screening compounds to identify agonists
XX and antagonists of E. faecalis protein activity
XX
XX Sequence 627 AA;
XX
XX Query Match 13.0%; Score 334; DB 2; Length 627;
XX Best Local Similarity 26.3%; Pred. No. 7.3e-12;
XX Matches 150; Conservative 79; Mismatches 223; Indels 118; Gaps 26;
XX
XX 3 KMIQSLVSLASLAFMAVSPVTPAFPA--ETGITVQDTQKGATYKAYKVFDAEIDNAN 60
```

```
Db 100 KOAVQSL-----TFCKPVAQGTTDANGVTVQLPKKONGKDAVYTIKEPKEGV 148
QY 61 VSDNKGCA--SYLLPOGKEAEYKASTDFNSLF-----TTTNGCRYYVTKKDTASANE- 112
Db 149 VAATNMVVAFPVYEMIKQTDGSKYKYTEELAVHHIYPKNVVANDGSLHVKVGTAE-NEG 207
QY 113 -----IATWA-----KSI SANTTPV--STVTESNND 136
Db 208 LNAEAFVUSKSESGPFTVKYIQGVKDGLYTWTDDKEQAKRFITGSKSYEIGENDFTEAEN- 266
QY 137 GTEVINVS--QYGYVYS--STVNGAVIMVTSVTP-----NATIHKEK--NTDATWGDG 184
Db 267 GTGELTVKNLEVGSYILEEVKAPNNAELIENQTKTPTTEANNQTPVEKTVKNDTSKVDK 326
QY 195 GGTVDQKTYSGVTVKTYIYKNAVNHGTE---KVYQVVIKDTMPSASVVD-LNEGS 239
Db 327 TTPSLDGDQVAIGEKIKYQISVNIPLGLIADKEGDANKYVKNLVKDDKDAALTFDNVTSGE 386
QY 240 YEVTTIDSGNITTLTGSEKATGKYNLLENNNFITII--PWAATNTPGTGNTONGANDD 297
Db 387 YAYALYDGDTVI-----APENYQVTEQANGFTVAVNPAYIPTLTPGG----- 428
QY 298 PFYKGINITVYTYGVLKSGAKPGSADLPENTNIATINPNTSNDPQGVTVRQGITIK 357
Db 429 -----TLKFVYFMHLNEKADPTKG---FKNEANVDNGHTDQDTPPTVEVVTGGRKFI 477
QY 358 KIDGSTKA--SIQGAIFVLKNA---TGQFLNFDNTNN-VEW-GTEANATEYTTGADGII 410
Db 478 KVDGIVTATQALAGASFVVRDQNSDTANYLKIDETTKAATVWKYKAEATFTTTADGLVD 537
QY 411 ITGLKEGYIYLVKKAIPGLYNLLDNSQKVIILGDGATDTNSDNLLVNPVTEN-NKGTBLP 469
Db 538 ITGLKGYIYLEETVAPDDYVLLTNRIEFVNVQSGYGT--ENLVSPEKVPNKHKGK-LP 594
QY 470 STGGIGTTFIYIIGAILVIGAGIVLVARRR 499
Db 595 STGGKGYIYVILGSGAVLLLIAGVVPARRK 624

RESULT 5
ABP43337
ID ABP43337 standard; protein; 627 AA.
XX
XX ABP43337;
XX
XX 05-AUG-2002 (first entry)
XX
XX E faecalis EF058 protein.
XX
XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.
XX
XX Enterococcus faecalis.
XX
XX US2002045737-A1.
XX
XX 18-APR-2002.
XX
XX 04-MAY-1998; 98US-00071035.
XX
XX 04-MAY-1998; 98US-00071035.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
XX WPI; 2002-425450/45.
XX
XX N-PSDB; ABN98093.
XX
XX New genes and polypeptides from Enterococcus faecalis, useful as vaccines
XX for preventing, treating or attenuating an infection caused by a member
XX of the Enterococcus genus in an animal, particularly E. faecalis.
```

PS Claim 9; Page 93; 255pp; English.

XX The present invention provides the protein and coding sequences of a

CC number of polypeptides from Enterococcus faecalis. The proteins can be

CC used as vaccines for preventing or attenuating an infection caused by a

CC member of the Enterococcus genus in an animal, particularly E. faecalis.

CC The polynucleotide is also useful for preventing or treating E. faecalis

CC infection. The present sequence is a protein of the invention

XX

SQ Sequence 627 AA;

Query Match 13.0%; Score 334; DB 5; Length 627;

Best Local Similarity 26.3%; Pred. No. 7.3e-12;

Matches 150; Conservative 79; Mismatches 223; Indels 118; Gaps 26;

QY 3 KMIQSLVASLAFGMVSPVTPFAFAA--ETGTTIVQDTQKGATYKAYKVFDAIDNAN 60

Db 100 KQAVQSL-----TPGKPVAGTTDANGNVTVQLPKKQNGDAVYTIKEPKEGV 148

QY 61 VSDSNKOGA--SYLIPQKGEAYKASTDFNSLF-----TTTTNGGRTVYTKDFTASANE- 112

Db 149 VAATNMVAVFPVYEMIKQTDGSKYKGTBELAVVHIYPKNVAVNDGSLHVKVGTAE-NEG 207

QY 113 -----IATWA-----KSIANTTPV--STVTESNND 136

Db 208 LNGAEFVISKSESGPGTVKVIQGVKDGLYTWTDKBQAKRFITGKSYEIGENDFTEAEN- 266

QY 137 GPEVINVS--QGYYYVVS--STVWNGAVIMVTSVTP-----NATIHEK--NTDATWGDG 184

Db 267 GTGELTVKNLGVSVILEBVKAPNAELIENQTKPTFTIEANNQTPVEKTVKNDTSKVDK 326

QY 185 GGTVDQKTYSGDVTVKYITTYKNAVNVHGTG-----KVYQYVIKDTPMSASVVD-LNEGS 239

Db 327 TTPSLDGKDVATGEIKYQISVNIPLGIADKEGDANKYVKFNLVDKHAALTFDNTVTSGE 386

QY 240 YEVTTIDSGNITLTQSEKATGKYNLLENNNTIIT--PWAATNPTGNTQNGANDD 297

Db 387 YAYALYDGDVTI-----APENYQVTEQANGFTVAVNPAYIPTLTPGG----- 428

QY 298 PFYKGINTITVYTGVLKSGAKPGSADLPENTNIATINPNTSDPDGQKVTVRDGOITIK 357

Db 429 -----TLKFVYFPHLNEKADPTKG-----FKNEAVNDGHTDDQTPPTVEVVTGKRFI 477

QY 358 KIDGSTKA--SLQGAIFVLKNA---TGQFLNFNDTN--VEW-GTEANATEYTTGADGIIT 410

Db 478 KVDGVTATQALAGASFVVRDQNSDTANYLKIDETTKAATWVKRAEATFTTTADGLVD 537

QY 411 ITGLKECTYLYVEKAPLGYNLNDSQKVLGDGATDITNSNLLVNPVEN-NKGTBLP 469

Db 538 ITGLKYGTYLYEETVAPDDYVLLTNRIEFVWNEQSYGTT--ENLVSPEKVPNKHKGTL-P 594

QY 470 STGGIGTTFYIIGAILVIGAGILVARRR 499

Db 595 STGGKGIYVLGSGAVLLIAGVYFARRRK 624

RESULT 6

ABU88365

ID ABU88365 standard; protein; 627 AA.

XX

AC ABU88365;

XX

DT 07-JUL-2003 (first entry)

XX

DE E. faecalis novel protein #109.

XX

XX Vaccine; endocarditis; bacteraemia; urinary tract infection; UTI;

KW intraabdominal infection; soft tissue infection; neonatal sepsis.

XX

OS Enterococcus faecalis.

XX

PN US2003017495-A1.

XX

PD 23-JAN-2003.

XX

PF 29-JUL-2002; 2002US-00206576.

XX

PR 06-MAY-1997; 97US-0044031P.

PR 16-MAY-1997; 97US-0046655P.

PR 14-NOV-1997; 97US-0066009P.

PR 04-MAY-1998; 98US-00071035.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;

XX WPI; 2003-416890/39.

DR N-PSDB; ACA88057.

DR

XX New nucleic acid molecules and polypeptides from Enterococcus faecalis,

PT useful as vaccines for preventing or attenuating an enterococcal

PT infection in an animal, or for identifying Enterococcus faecalis in

PT biological samples.

XX

PS Claim 12; Page; 40pp; English.

XX

CC The invention relates to a new isolated nucleic acid molecule comprising

CC a polynucleotide isolated from Enterococcus faecalis appearing as

CC AC87949-ACA88196 (or sequences complementary to them or 95% identical to

CC them). Also included are the proteins encoded by the above nucleic acids,

CC making a recombinant vector (comprising inserting the isolated nucleic

CC acid molecule cited above into a vector), a host cell comprising the

CC vector (used to produce the protein), an isolated antibody specific for

CC the polypeptides, a hybridoma that produces the antibody, an isolated

CC polypeptide antigen comprising an amino acid sequence of an Enterococcus

CC faecalis epitope listed in the specification, a vaccine comprising one or

CC more E. faecalis polypeptides (and a pharmaceutical diluent, carrier or

CC excipient) where the polypeptide elicits protective antibodies in an

CC animal to a member of the genus Enterococcus; preventing or attenuating

CC an infection caused by a member of the genus Enterococcus in an animal

CC comprising administering to the animal the polypeptide and detecting

CC Enterococcus nucleic acids in a biological sample. The E. faecalis

CC nucleic acid molecules and polypeptides are useful as vaccines for

CC preventing or attenuating an enterococcal infection in an animal (e.g.

CC endocarditis, bacteraemia, urinary tract infection in a human).

CC infection, soft tissue infection and neonatal sepsis). The polypeptides

CC are also useful for detecting Enterococcus aureus in immunoassays, as

CC epitope tags, as molecular weight markers, or for generating antibodies

CC that specifically bind E. faecalis polypeptides. The nucleic acid

CC molecules are also useful as probes for gene mapping, or for identifying

CC E. faecalis in biological samples. The kit and methods are useful for

CC detecting Enterococcus antibodies or nucleic acid molecules in a

CC biological sample. The present sequence is a novel E. faecalis

CC polypeptide of the invention. Note: The sequence data for this patent did

CC not form part of the printed specification, but was obtained in

CC electronic format directly from the USPTO at

CC seqdata.uspto.gov/sequence.html?docID=20030017495

XX

SQ Sequence 627 AA;

Query Match 13.0%; Score 334; DB 6; Length 627;

Best Local Similarity 26.3%; Pred. No. 7.3e-12;

Matches 150; Conservative 79; Mismatches 223; Indels 118; Gaps 26;

QY 3 KMIQSLVASLAFGMVSPVTPFAFAA--ETGTTIVQDTQKGATYKAYKVFDAIDNAN 60

Db 100 KQAVQSL-----TPGKPVAGTTDANGNVTVQLPKKQNGDAVYTIKEPKEGV 148

QY 61 VSDSNKOGA--SYLIPQKGEAYKASTDFNSLF-----TTTTNGGRTVYTKDFTASANE- 112

Db 149 VAATNMVAVFPVYEMIKQTDGSKYKGTBELAVVHIYPKNVAVNDGSLHVKVGTAE-NEG 207

QY 113 -----IATWA-----KSIANTTPV--STVTESNND 136

Db 208 LNGAEFVISKSESGPGTVKVIQGVKDGLYTWTDKBQAKRFITGKSYEIGENDFTEAEN- 266



PF 30-JUN-1998; 98US-00107532.  
PR 02-JUL-1997; 97US-0051571P.  
PR 14-MAY-1998; 98US-0085598P.  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Doucette-Stamm LA, Bush D;  
DR WPI; 2003-799836/75.  
DR N-PSDB; ADC93971.  
XX New isolated nucleic acid derived from Enterococcus faecium encoding an  
PT Enterococcus faecium polypeptide useful for detection, prevention and  
PT treatment of a pathological condition resulting from a bacterial  
PT infection.  
XX  
PS Example 1; SEQ ID NO 7252; 243pp; English.  
XX  
CC The invention relates to an isolated nucleic acid derived from  
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridising to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to a  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids is useful for diagnosing pathological conditions  
CC resulting from E. faecium bacterial infection (e.g. urinary tract  
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
CC infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of Candida albicans -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating Enterococcus faecium infections. The present sequence represents  
CC one if the disclosed E. faecium proteins.  
XX  
SQ Sequence 664 AA;  
  
Query Match 12.3%; Score 318; DB 7; Length 664;  
Best Local Similarity 24.2%; Pred. No. 7.2e-11;  
Matches 141; Conservative 91; Mismatches 206; Indels 144; Gaps 28;  
  
QY 3 KMIQSLVSLAFGMVSPVTPFAA--ETGTTV---QDTQKATYKAYKVFDAAI 56  
DB 137 KQAVQSL-----TPGTPVAGTTDAGNVTLSPKKQNGKDAVY----- 175  
  
QY 57 DNANVSDSNKOGAS-----YLIQKGEABYKAST---DPNSLF--TTTNGGRT 100  
DB 176 ---TIKEBPKDGVSAAANVLAPVYEMIKQADGSYKYGTBELDTHLYPKNTVGNQDGL 232  
  
QY 101 VYTKDTSANRATWAKSISANT--TP-----VSTVTE----- 132  
DB 233 KVTIKGTAE-NEALNGAFELISKEGTSVKYKIQSVTDGLYTTWTDQKAKHFTGHSY 291  
  
QY 133 --SNND-----GTEVINVSQYGYVYVS--STVNGGAVIMVTSVTPNATHEKNT--- 177  
DB 292 DIGNNDFAEASIEKQLIVNHLVEVKYNLEKVPADNAEMIEKQITPFEILANSQTPVE 351  
  
QY 178 -----DATWGGGGKTVQKTSVSGDTVKYITTYK---NAVNHGTEKY-QYVTKDTP 228  
DB 352 KTRKNDTSKVDKTRTPQLNGKDAVIGEKIQEISVNPILGDIADKEGTQNKYTFKLDTHD 411  
  
QY 229 SASVVDL-NEGSVEVITDGSNITTLTQSGSEKATGKYNLLENNNFTITP--WAATNT 285  
DB 412 AALTENDSSGYALVDYDNGKEIDPV-----NYSVTEQDTGTVSDPNYPSLT 462  
  
QY 286 PTGNTQNGANDFFYKGINITITVYTVGLKSGAKPGSADLPENTINATINPNSTNDPDQ 345  
DB 463 FGG-----TLKFVYVYMLNEKADPTKG-----FSNQAVDNGHTNDQTPP 502  
  
QY 346 KVTVRDGGQITIKIDG--STKASLQGAIFVLKNA---TGQFLNFN-DTNNVFN-GTEANA 398

DB 503 SVDVVTGGRKRVKVDGVDVTSQTLGAEFVVRDQSDTAKYLSIDPSTKAVSWSAKESA 562  
QY 399 TEYTTGADGIITITGLKEGTYLVKPKAPLGNLLDNSQKVLGDBGATDTTNSDNLVNP 458  
DB 563 TVFTTTSNGLIDVTLGKYGTYLLEETKAPKPYVPLTRVATIDEQSVVTAGQ---LISP 619  
QY 459 --TVENNKGTLPSTGGTGTTFYIIGAILVIGAGIVLVARR 498  
DB 620 EKIPNKHGT-LPSTGGKGIYVIGAGVWLLLIAGLYFARK 660  
  
RESULT 9  
ABP27284  
ID ABP27284 standard; protein; 705 AA.  
XX  
AC ABP27284;  
XX  
DT 02-JUL-2002 (first entry)  
XX  
DE Streptococcus polypeptide SEQ ID NO 3744.  
XX  
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
OS Streptococcus agalactiae.  
XX  
PN WO200234771-A2.  
XX  
PD 02-MAY-2002.  
XX  
PF 29-OCT-2001; 2001WO-GB004789.  
XX  
PR 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
DR WPI; 2002-352536/38.  
DR N-PSDB; ABN67915.  
XX  
PT New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
PS Claim 1; Page 3529; 4525pp; English.  
XX  
CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
XX Streptococcus proteins  
SQ Sequence 705 AA;

Query Match	10.8%; Score 279.5; DB 5; Length 705;
Best Local Similarity	24.3%; Pred. No. 1.6e-08;
Matches 149; Conservative	76; Mismatches 213; Indels 175; Gaps 30;
QY	19 AVSPVTP-----IAFAAE--TGTITVQDTQKGATY-----KAYKVFDAEIDNANV 61
Db	132 AVSGLTPEGDTGLVFNTKGLGKEFKIVEKSKSTYNNNGSLLAASKAVPV-----NITL 185
QY	62 SDSNKGDA---SYLIPQGEA-----EYKASTDFNSLFTTTTNGRVTYTKD 106
Db	186 PLVNEGVVADAHVYPKNTKEPEIDKNFAKTNLDLTALTDVNRLLTAGANYG-NYARDKA 244
QY	107 TASAN-----ETAT-----WAKSISANTTPVSTVTESNNDGTGV----- 140
Db	245 TATAEIGKVPVEVKTKIHGSKYENLVWTDIMSGLTWSVLSKASGTTTFPAKQD 304
QY	141 -INVSQYGY-----YVVSSTVNNAGVIMVTSVTPNATIEHK 175
Db	305 ELSIDARGFTLKFTADGLGKLEKAAKTADIEFTLTYSATVNGQAI-----DNPSN 356
QY	176 NYDATWGDGGKTVDO--KTVSGDVTKYITTYKNAVNYHGTCKYQYVVKD----- 225
Db	357 DIKLSYGNKPGKDLTELPVTPSKGE-VTVAKTWSGDGIAPDGVNVVYTLKDKDKTVASVSL 415
QY	226 TMPSASVVDLNEG-SYEVTITDGSNITTLTQ-----SEKATGKYNLLE-ENNFT 275
Db	416 TKTSKGTIDLNGIKFEV-----SGNPSGKFTGLENKSYMISERVSGYSAINLENGKVT 470
QY	276 ITIPWAATN--TPGTNQ-----NGANDFFYKGINIITVVTG---VLKGA 318
Db	471 ITNTKDSNPTPLNPTPEKPVETHGKFKVKTNEQDRL--AGAQFVVKNSAGKYLAKADQ 528
QY	319 KPGSADLPE-----NTNIATINPNTSDPDQKVTYRDGQITIKKIDGSTKASLOQAI 373
Db	529 SEGQKTAAKIALDEAIAAYNKLSDQKGEK-----GITAKELIKTQADYDAAFTE 582
QY	374 LKNATGQFLNFNDTNVWEGTE-ANATEYTTGADGIITITGLKEGYTYLVEKAPLGYNL 432
Db	583 ARTA-----YEWITDKARAITYTSNQDQGFVETGLADGTYNLEETLAPAGFAK 630
QY	433 LDNSQKVLGDGATDTTNSDNLVNPT-----VENNKGTLPSTGGICTTIFYIIGAIL 486
Db	631 LAGNIKPVVNGSYITGGNIDYVANSNQDKATRVENKKVT-IPQTGGIGITLFTIIGLSI 689
QY	487 VICAGIVLVARRR 499
Db	690 MLGA-VVIMKREQ 701
RESULT 10	
AAV00119	standard; protein; 560 AA.
XX	AAV00119;
AC	AAV00119;
XX	20-APR-1999 (first entry)
DT	
DE	Enterococcus faecalis antigenic polypeptide fragment EF058.
XX	
KW	Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW	detection; attenuation; antigenic.
XX	
OS	Enterococcus faecalis.
XX	
FN	WO9850554-A2.
XX	
PD	12-NOV-1998.
XX	
EF	04-MAY-1998; 98WO-US008959.
XX	
PR	06-MAY-1997; 97US-0044031P.
PR	16-MAY-1997; 97US-0046655P.
PR	14-NOV-1997; 97US-0066009P.

XX	(HUMA-) HUMAN GENOME SCI INC.
PA	Kunsch CA, Choi GH, Bailey C, Hromockyj A;
XX	
PI	WPI; 1999-070095/06.
XX	N-PSDB; AAX20109.
DR	
DR	
XX	New isolated Enterococcus faecalis polynucleotides - used to develop
PT	products for the detection of Enterococcus and for use in vaccines for
PT	prevention or attenuation of Enterococcus infection.
XX	
XX	Claim 9; Page 134; 301pp; English.
XX	
CC	The present sequence represents an antigenic polypeptide fragment
CC	isolated from Enterococcus faecalis. The present invention describes
CC	genes, proteins and antigenic polypeptides isolated from E. faecalis. The
CC	proteins can be used in vaccines for preventing or attenuating an
CC	infection caused by a member of the Enterococcus genus in a sample. The
CC	can also be used for detecting Enterococcus antibodies in a sample. The
CC	nucleotide sequences can be used for detecting Enterococcus nucleic
CC	acids. Products from the present invention can also be used for screening
CC	compounds to identify agonists and antagonists of E. faecalis protein
CC	activity
XX	
QY	Sequence 560 AA;
Db	
QY	Query Match
Db	Best Local Similarity
QY	Matches 136; Conservative
Db	72; Mismatches 212; Indels 117; Gaps 25;
QY	3 KKMIOQLLVASLAFGMVSPVPTIAFAA--ETGTTITVQDTQKGATYKAYKVFDAEIDNAN 60
Db	68 KQAVQSL-----TPGKFAVAGTTDANGNVTVQLPKKQNGKDAVYTIKEPEKGV 116
QY	61 VDSNKGDA--SYLIPQGEAEYKASTDFNSLF-----TTTTNGRVTYTKOTASANE- 112
Db	117 VAATNMVAFVYEMIKQTDGSKYKGTTELAVHIIYPKNVANDGSLHVKVGTAE-NEG 175
QY	113 -----IATWA-----KSIANTTPV--STVTESNND 136
Db	176 LNGAEFVLSKSESGPGTVKYIQGVKGLYTWTTDKQAKRFTGKSYEIGENDFTEAEN- 234
QY	137 GTEVINVS--QYGYVVS--STVNNGAVIMVTSVTP-----NATIEHK--NTDATWGDG 184
Db	235 GTGELTVKNLEVGSYILEEVKAPNNAELIENQTKTPTTEIANNQTPVEKTVKNDTSKVDK 294
QY	185 GKTVDOKTSYSGDTVKYITTYKNAVNYHGT-----KVQYVVKIDTMPSASVVD-LNEGS 239
Db	295 TTPSLDGKDVAGIEKIKYQISVNIPLGIADREGDKANKVKNLVDKHDALTFDNTVSGE 354
QY	240 YEVTTITDGSNITTLTQSGSEKATGKYNLLENNFTITI--PWAATNTPTGNTONGANDD 297
Db	355 YAYALYDGDTVI-----APENYQVTEQANGFTVAVNPAYIPTLTPGG----- 396
QY	298 FFYKGINITVYTGVLKSGAKPGSADLPENTNIIATINPNTSDPGQKTVYRDGQITIK 357
Db	397 -----TLKFVTFMHLNEKADPTKG-----FKNEANVDNGHTDDQTPPTVEVVTGGRKFI 445
QY	358 KIDGSTKA--SLQGAIFVLKNA---TGQFINFNDTNN-VIEW-GTEANATEYTTGADGIIT 410
Db	446 KVDGVDVTAQALAGASFVVRDQNSDTANYLKIDETKKAATWVKTKAEATFTTTADGLVD 505
QY	411 ITGLKEGYTYLVEKAPLGYNLNDSQKVLGDGATDTTNSNLLVNPTVEN-NKGT 466
Db	506 ITGLKGYTYLLEETVAPDDYVLLTNRIEFVNVNEQSYGTT--ENLVSPEKVPKNKHGT 560
RESULT 11	
ID	ABP43338 standard; protein; 560 AA.
XX	ABP43338
AC	ABP43338;







XX AC ABU00821;  
 XX DT 23-OCT-2003 (revised)  
 XX DT 11-FEB-2003 (first entry)  
 XX DE S. pneumoniae type 4 strain protein from coding region #389.  
 XX DE Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.  
 XX OS Streptococcus pneumoniae; type 4 strain.  
 XX PN WO200277021-A2.  
 XX PD 03-OCT-2002.  
 XX PF 27-MAR-2002; 2002WO-IB002163.  
 XX PR 27-MAR-2001; 2001GB-00007658.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Massignani V, Tettelin H, Fraser C;  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ABX06101.  
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 PS Claim 1; SEQ ID NO 778; 56pp; English.  
 XX The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target  
 CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIFO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX SQ Sequence 665 AA;

Query Match 10.3%; Score 265.5; DB 6; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 1e-07;  
 Matches 164; Conservative 63; Mismatches 227; Indels 219; Gaps 33;

QY 9 LLVASLAFGMAVSPVPIAFAAETGTITVQDTQKGAATYKAYKVFDAEIDNANVSDSNKDG 68  
 |||||  
 Db 15 LLTASSLPSAAT-----VFAAGTTTTSVTHKLLATDGDMDKIANELETGNYA--GNKVG 67  
 |||||  
 QY 69 ASYLIPQGEAEYKASTDFNSLFTTTN-----GGRTVTVTKKDTASANEIATWAKSISAN 123  
 |||||  
 Db 68 ---VLP-----ANAKEIAGVMFVNTNNEIIDENGQTLGVNDP-----QTFKLSGAMP 114  
 |||||  
 QY 124 TTPVSTVTESNNDGTEVINV--SOGVYVVS-----TVNNGAVIMVTSVTP----- 168  
 |||||  
 Db 115 ATAMKLTAEAGAKENTANLPAKYKIYEHLSLTVYGEDGATLTGSKAVFIEIPLND 174  
 |||||  
 QY 169 --NATIEKNTDA--TWGDDGGKT-----VDQKT-----YSGDVTVKYTIYK--NAVNY 212  
 |||||  
 Db 175 VVDAHVPKNTAKPKIDKDFKXANPDTPRVDKDTPVNHQGVGVVEIYTKIPALANY 234  
 |||||  
 QY 213 -----HGETKYY-----QYVIKDT-----MPSASVDLNEG-- 238  
 |||||  
 Db 235 ATANWSDRMTEGLAFNKGTVKVTVDVALEAGDYALTAVATGFDLKLTDAGLAKVNDQNA 294  
 |||||  
 QY 239 -----SYEVTITDGS-----GNITLTQSEKATG---KYNLLENNNFTITPWA-A 282  
 |||||  
 Db 295 EKTVKITYSATLNDKAIVEVPESNDVTFNNGNPDHGNTPKPNKDNENGDLTLTKTWDA 354  
 |||||  
 QY 283 TNTPTG-----NTQNGANDDFFYKGINTI-----TVTYTGVLKS-----G 317  
 |||||  
 Db 355 TGAPIPAGAEATFDLVNAQTG-----KVQTVTLTTDKNTVTNGLDKNTEYKFVRS 407  
 |||||  
 QY 318 AKPSADLPENTNATI-----NPNTSNDPQKQKVTVRDQGITIKKIDGSTKASLQGA 371  
 |||||  
 Db 408 IKGYSADYQETITTAAGEIAVKMKDKENPKPLDPTBPVKVTVYKGFVKVNDKDNR--LAGAE 465  
 |||||  
 QY 372 FVLKNA--TGQFL-----  
 |||||  
 Db 466 FVIANADNAGYLARKADKVSQBEKQLVVTTYKDALDRAVAAYNALTAQQOQTQOEKVKDK 525  
 |||||  
 QY 383 ---NFN-----DTNNVEWGTG---ANATEYTTGADGIIITGLKEGTYYLVEKKAPLGYN 431  
 |||||  
 Db 526 AQAAYNAAVIAANNAFVADKDNENNVKLVSDAQGRFEITGLLAGTYYLEETKQAGYA 585  
 |||||  
 QY 432 LLDNSQV-----ILGDGATDTTNSDNLVNPTVNNKGTBLPSTGGIGTTIFYIIG 483  
 |||||  
 Db 586 LLTSRQRFVATSYATSGQIEYTAGSGK--DDATKVNKKITIPQTGGIGTIIFAVAG 643  
 |||||  
 QY 484 AILVIGAGIVLVA 496  
 |||||  
 Db 644 AAIM---GIAVYA 653

## RESULT 15

AAY81627

ID AAY81627 standard; protein; 666 AA.

XX AC AAY81627;

XX DT 24-MAY-2000 (first entry)

XX DE Streptococcus pneumoniae type 4 protein sequence #127.

XX KW Streptococcus pneumoniae; vaccine; screening; protein antigen;  
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
 XX KW pneumococcal disease.  
 XX OS Streptococcus pneumoniae.  
 XX PN WO200006737-A2.  
 XX PD 10-FEB-2000.  
 XX PF 27-JUL-1999; 99WO-GB002451.  
 XX PR 27-JUL-1998; 98GB-00016337.

PR 19-MAR-1999; 99US-0125164P.  
XX (MICR-) MICROBIAL TECHNIKS LTD.  
XX  
XX Gilbert CFG, Hansbro PM;  
XX WPI; 2000-195300/17.  
XX  
XX New Streptococcal protein, useful as a vaccine, for diagnosis of  
PT pneumococcal diseases and for screening agents capable of antagonizing or  
PT inhibiting expression of the protein.  
XX  
XX Claim 1; Page 85; 108pp; English.  
XX  
XX AA051501 to AA051679 represent specifically claimed protein sequences  
CC isolated from Streptococcus pneumoniae. AA051501 to AA051590 represent  
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.  
CC The sequences have antibacterial and antiinflammatory properties. The  
CC protein sequences, and fragments of them, are useful as immunogens and/or  
CC antigens. The nucleotide sequences can be used in vaccines and in  
CC diagnostic assays. The proteins and nucleotides can be useful for the  
CC detection and diagnosis of S. pneumoniae. The protein sequences are also  
CC useful for screening an agent capable of antagonising, inhibiting or  
CC interfering with the function or expression of the proteins in which the  
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection  
CC and meningitis. AA051591 to AA051614 represent primers used in the  
CC exemplification of the present invention  
XX  
XX Sequence 666 AA;  
SQ  
  
Query Match 10.3%; Score 265.5; DB 3; Length 666;  
Best Local Similarity 24.4%; Pred. No. 1e-07;  
Matches 164; Conservative 63; Mismatches 227; Indels 219; Gaps 33;  
  
QY 9 LTVASLAFGNAVSPVTFIAFAETGTYVDQTKGATYKAYKVFDAEIDNANVSDNKG 68  
DB 15 LTFASLFSAA-----VFAAGTTTSVTHKLLATDGMCKIANELETGNYA-GNKVG 67  
  
QY 69 ASYLIPQKGEAEYKASTDFNSLFTTTN-----GGRTYVTKKDTASANEIATWAKSISAN 123  
DB 68 ---VLP-----ANAKEIAGVMFVNTNNEIIDENGQTLGNIDP-----QIFKLSGAMP 114  
  
QY 124 TTPVSTVTESNNDGTEVINV--SQYGYVVS-----TVNGAVIMVTSVTP----- 168  
DB 115 ATAMKLLTEAGKAFNTANLPAKAYIETHSLSTYVGEDGATLTGSKAVPIEIELPLND 174  
  
QY 169 --NATHEKNTDA--TWGPGGCKT-----VDQKT--YSGVDTVKYITTYK--NAVNY 212  
DB 175 VDDAHVTPKTEAKPKIDKDFGKANPDPTRVDKTPVNHQGVGVVEYELVTKIPALANY 234  
  
QY 213 -----HGTEKVY-----QYVIKDT-----MPSASVVDLNEG-- 238  
DB 235 ATANWSRMTGLEAFNKGTVKTVTDVDALEAGDYALTEVATGFDLKLTDAGLAKVNDQNA 294  
  
QY 239 -----SYEVTITDGS-----GNITTLFGSEKATG---KYNLLENNTTITIPWA-A 282  
DB 295 EKTVKITYSATLNDKAIVEVPESNDVTFNNGNPDHGNTPKPNKPNENGDLTLTKTWVDA 354  
  
QY 283 TNPPTG-----NTQNGANDFFYKGINTI-----TVTYTGLVLS-----G 317  
DB 355 TGAPIPAGAEATFDLNAQTG-----KVQIVTLTDKNIVTVNGLDKNTYKVFERS 407  
  
QY 318 AKPGSADLPENTNIAT-----NPNTSNDPQKVTVRDQGITIKKIDGSTKASIQGAI 371  
DB 408 IKGYADYQBITTAGETAVKNWKENPKPLDPTPEKVVTVYKGFVKVNDKDNK--LAGAE 465  
  
QY 372 FVLKNA--TCQFL----- 382  
DB 466 FVIANADNAGQYLARKADKVSQBEKQLVWTTKDALDRAVAAYNALTAQOQTOQEKVKDK 525  
  
QY 383 ---NFN-----DTNNVEWGPTE---ANATEVTTGADGIITITGLKEGTYVLVEKKAPLGYN 431  
DB 526 AQAYNAAVTAANNAFEWADKDNENVVVLVSDAQGFETITGLLAGTYLLEETKQAGYA 585

QY 432 LLDNSQKV-----ILGDGATDTTNSDNLNVNFTVENNKGTETLPSTGGTITFIYIG 483  
DB 586 LITSRQKEVATSYSATGQIEYTAGSGK--DDATKVVNKKITIPQTGGITIFAVAG 643  
  
QY 484 AILVIGAGILVA 496  
DB 644 AAIM---GIAVYA 653  
  
RESULT 16  
ID ABB54071 standard; protein; 614 AA.  
XX  
XX ABB54071;  
XX AC ABB54071;  
XX 29-AUG-2003 (revised)  
DT 16-MAY-2002 (first entry)  
XX  
XX Lactococcus lactis protein yhgE.  
XX Lactococcus lactis protein yhgE.  
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
XX Lactococcus lactis; IL1403.  
XX Lactococcus lactis.  
PN FR2807446-A1.  
XX  
XX 12-OCT-2001.  
XX  
XX 11-APR-2000; 2000FR-00004630.  
XX  
XX 11-APR-2000; 2000FR-00004630.  
XX  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
XX WPI; 2002-043418/06.  
XX  
XX New nucleotide sequence useful in the identification of Lactococcus  
PT lactis and related species.  
XX  
XX Claim 6; SEQ ID NO 773; 2504pp; French.  
XX  
XX The present invention is related to a Lactococcus lactis nucleotide  
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic  
CC acid sequence is useful in the detection and/or amplification of nucleic  
CC acid sequence, particularly to identify Lactococcus lactis or related  
CC species. The proteins of the invention are useful for the biosynthesis or  
CC biodegradation of a composition of interest. The invention helps research  
CC in lactic bacteria, particularly useful in the production of yogurt and  
CC cheese. Note: The sequence data for this patent is based on equivalent  
CC patent WO200177334 (published 18-OCT-2001) which is available in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 29-AUG-2003 to  
CC standardise OS field)  
XX  
XX Sequence 614 AA;  
  
Query Match 10.0%; Score 258.5; DB 5; Length 614;  
Best Local Similarity 24.5%; Pred. No. 2.4e-07;  
Matches 145; Conservative 57; Mismatches 184; Indels 207; Gaps 31;  
  
QY 51 VFDAEIDN-----ANVSDSNKDGASYLIPOGKEAEYKASTDFNSLFTTTNGRTVYT 103  
DB 75 VNSGIANNQTCGAGNTGSTRKNDGSGQTTLEGSTPSTMANVTFSATKYVGT-GVPTGVT 133  
  
QY 104 K---KDTASANEIATWAKSISANT-----TPVSTVTESNNDGTEVINVSQ----- 145  
DB 134 DPFSTDTTTSVAPVTTDASGLADFTGLTGYLFLHQVTTTNGITTVGDFIVQVSHEDSOAG 193  
  
QY 146 ----YGYVVSSTVNAGVIMVTSVT-----PNATIEH-----KNITD--- 178

```
Db 194 IVNVPKLDMSSAGLG-----TSATTNADNPNFGQTPNQIANPNATGNSDOTLTNTDNN 248
Qy 179 -----ATWGGGGKTVQDKYSGVDFVKYITYKNV-----NYHGTEKYV-QXV 222
Db 249 AGNENLANGTWTNGSDNQ-NTTTAAAGNTVNNV---NTVPDSSQTNNNGTGTGTGYI 304
Qy 223 IKDTP-----SASVVDLNEGS-----YEVTTDSCGNT--TTLTQG 257
Db 305 VTDQLPNNLVNSTVTVSVIVNVNNGSGTKVGTLTPTTDTYIT-NDGNGKIVVTLTTAG 363
Qy 258 SEKATGKYNLLBENNFTIIPWAATNPTGNTONGANDDPFYKGINITITVYTGVLKSG 317
Db 364 QHRAASL--LGSADGALNIIIP-STVKSAGSATSAT-----TITNAY----- 405
Qy 318 AKPGSADLPENTNIATINPNTSNDPQGVTVRDGQITIKKIDGSTKASLQGAIFVLKNA 377
Db 406 ----GADLTSTTAV-----KSTLVGGLMTKTDASTNAALAGATTVVRA 447
Qy 378 TCG-----PLNFND-----TNVWEGTEANATE-----YTTGAD 406
Db 448 DNKEDAQDFEANAAYFNNSASGGIVTNLTSSKAAPVTGDTSGNANTSATAPVTFITGKD 507
Qy 407 GIITITGLK-----EGT-----YILVEKKAPLGYNL-----LDNSQKVLGDGATDT 448
Db 508 GIATFNGNLVNNDGNTWTHYILVEVAAPTGYQLPSVTTAANTGTAVTASTAPAAITD 567
Qy 449 TNSDNLVNPTVNNKGTLPSTGGIGTTFYIIGAILVIG--AGIVLVARRR 499
Db 568 -----TITNKKFPALPFTGGG-----LAGIATVSGVIAFAIKR 604

RESULT 17
ID ABP29802 standard; protein; 554 AA.
XX AC ABP29802;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 8780.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB004789.
XX PR 27-OCT-2000; 2000GB-00026333.
XX PR 24-NOV-2000; 2000GB-00028727.
XX PR 07-MAR-2001; 2001GB-00005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN70433.
XX XX
XX PT New Streptococcus protein for the treatment or prevention of infection or
XX PT disease caused by Streptococcus bacteria, such as meningitis, and for
XX PT detecting a compound that binds to the protein.
XX XX
XX PS Claim 1; Page 3995; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
```

```
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX Sequence 554 AA;
SQ
Query Match 9.5%; Score 244.5; DB 5; Length 554;
Beat Local Similarity 25.4%; Pred. No. 1.5e-06;
Matches 151; Conservative 61; Mismatches 218; Indels 165; Gaps 31;
Qy 1 MKKKMIQSLLVASLAFGNVSPVTPPIAFAAETGTV-----QDTQKGAIFYKAVK--- 51
Db 3 LSKKLLFSAALVTWAGSTVEPVAQFA---TGMISIVRAAEVSQERPAKTTVNIYKLOAD 58
Qy 52 -FDAEIDNANVSDSNKDG--ASYLIPQKGEAEYKASTDFNSLFTTTNGGRTYVTKDITA 108
Db 59 SYKSEI-TSNGGIENKDEVISNYAKLGDNVKLGQGVF-----KRYKVKTD-I 105
Qy 109 SANEIATWAKSISANTTPVSTVTE-----SNNDGTEV---INVQYGYVYSSVTVNN 157
Db 106 SVDELKK-LTTVEAADAKVGTILEEGVSLPQKTAQGLVVDALDSKSNRVLYVEDLKNS 164
Qy 158 GAVIMVTSVTP-----NAT-----IHEKN--TDATWGGGGKTVDQKTSVGD 198
Db 165 PSNITKAYAVPFVLELPVANSTGTGFLSEINIYPRNVVTDKPKDKVKLQG-----D 218
Qy 199 TVKTYITYKNAVNVHGTGKQYQVYIKDTMPSASVVDLNEGSYE-VTITDGSNITTLTQG 257
Db 219 DAGYTI-----GEE--PKWFLKSTIPA-----NLGDYKFEITDKFADGLTY--- 258
Qy 258 SEKATGKYNL-----LEENNNFTITIPWAATNPTGNTONGANDDP-----FYKGI 303
Db 259 --KSVGKIKIGSKTLNRDEHYTI-----DEPTVDNQNTLKITFKPEKFEIABELLKGM 309
Qy 304 NTITVYTYGVLKSGAKPGSADLPENTNIATI-----NPN 337
Db 310 -TLVKNQDALKATANTDDAAFLPIPVASTINEKAVLGKAIENTFELQYDHTPDKADNPK 368
Qy 338 TSNDPQGVTVRDGQITIKKIDGSTKASLQGAIFVLKNATGQFLNFDNTNNVWEGTEAN 397
Db 369 PSN--PPRKPEVHTGGKRFVKDSTETQTLGSAFEDLLASDGTAVKWTDA-LIKANTKN 425
Qy 398 --ATEYTTG-----ADGIITITGL-----KEG---TYLVEKKAPLGYNLNDSQ 437
Db 426 VIAGEAVTGPILKLSHTDGTGFEIKGLAYAVDANAEGTAVTYKLKETRAPEGVVPDKEI 485
Qy 438 KVIILGDGA-----TDTTNSDNLVNPTVENKGTLPSTGGIGTTFYIIGAILV 487
Db 486 EFTVSQTSYNTKPTDITVDSADATPDPTIKNNKRPSPINTGGTGAIFVAIGAAMV 540

RESULT 18
ABP27373
ID ABP27373 standard; protein; 581 AA.
XX AC ABP27373;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 3922.
```

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX Streptococcus agalactiae.  
OS  
XX WO200234771-A2.  
XX PN  
XX PD  
XX 02-MAY-2002.  
XX 29-OCT-2001; 2001WO-GB004789.  
XX PF  
XX 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX PR  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
PA Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX WPI: 2002-352536/38.  
XX DR N-PSDB; ABN68004.  
XX DR  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX PT  
XX Claim 1; Page 3546; 4525pp; English.  
XX PS  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX SQ  
Sequence 581 AA;  
Query Match 9.5%; Score 244.5; DB 5; Length 581;  
Best Local Similarity 25.4%; Pred. No. 1.6e-06;  
Matches 151; Conservative 61; Mismatches 218; Indels 165; Gaps 31;  
QY 1 MKKQMIQSLVLASLAFMGAVSPVTPFAAFTGTTV-----QDTQKATYKAYKV--- 51  
DB 30 LSKKLLFSAALVLTWAGSTVEPVQAFA---TGMISVRAAEVSOERPAKTVNIYKLDAD 85  
QY 52 -FDAIDNANVDSNKG--ASYLIPQKEAEYKASTDFNSLFTTTTNGRTYVYKKDTA 108  
DB 86 SYKSEI-TSNGGIENKGEVINSYAKLGDNVKGLQGQVF-----KRYKVKTD-I 132  
QY 109 SANETATWAKSISANTTPVSTVTE-----SNNDGTEV---INVSQYGYVYSSTVNN 157  
DB 133 SVDELKK-LTTVEARADAKVGIILEEGVSLPQKTAQGLVVDALDSKNVRVLYVEDLKN 191  
QY 158 GAVIMVTSVTP-----NAT-----IHEKN---TDATWGDGGKTVQDKTYSVD 198  
DB 192 PSNITKAVAPVPLEPVLNANSTGTGFLSEINIPKNNVTVDEPKTDKQVKLQ-----D 245  
QY 199 TVKYTYITKAVNNYHGTKEVQYVIKDTMPSASVVDLNEGSVE-VTITDGSNTITLTQ 257

DB 246 DAGYTI-----GEE--FKWFLKSTIPA-----NLGDYKFEITDFADGLTY--- 285  
QY 258 SEKATGKYNL-----LENNFTTITIPAAATNTPTGNTONGANDDF-----FYKGI 303  
DB 286 --KSVGKIKIGSKTLNRDEHYTI-----DEPTVDNQNTLKITFKPEKFEIAELLKGM 336  
QY 304 NTITVTVTGVLKSGAKPGSADLPENTNIATI-----NPN 337  
DB 337 -TLVKNQDALDKATANTDDAAFLPIPVASTINEKAVLGKALIENTFELQYDHTDPKADNPK 395  
QY 338 TSNDPQOKVTVRDGQITIKIDGSTKASLOGAIFVLKATNQGLNFNDNTNNVWGTEAN 397  
DB 396 PSN--PPRKPEVHTGGRFVKKSTETQTLGGABFDLLASDGTAVKWTDA-LIKANTNKN 452  
QY 398 --ATEYITG-----ADGIITITGL-----KEG---TYLVVEKKAPLGYNLLDNSQ 437  
DB 453 YIAGEAVTGQPIKLKSHDTGTFEIKGLAYAVDANAEGTAVTYKLTKEPGEYVDPKEI 512  
QY 438 KVIILGDGA-----TDTTNSDNLVNPTVENNKGTELPSTGIGTTFIVIGAILV 487  
DB 513 EFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPTNGTGGTGAIFVAIGAAYM 567  
RESULT 19  
ADC95757  
ID ADC95757 standard; protein; 341 AA.  
XX AC ADC95757;  
XX DT 01-JAN-2004 (first entry)  
XX DE E. faecium protein sequence SEQ ID 5384.  
KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
KW abdominal-pelvic infection.  
XX OS Enterococcus faecium.  
PN US6583275-B1.  
XX PD 24-JUN-2003.  
XX PF 30-JUN-1998; 98US-00107532.  
XX PR 02-JUL-1997; 97US-0051571P.  
PR 14-MAY-1998; 98US-0085598P.  
XX PA (GENO-) GENOME THERAPEUTICS CORP.  
XX PI Doucette-Stamm LA, Bush D;  
XX WPI; 2003-799836/75.  
DR N-PSDB; ADC92103.  
XX New isolated nucleic acid derived from Enterococcus faecium encoding an  
PT Enterococcus faecium polypeptide useful for detection, prevention and  
PT treatment of a pathological condition resulting from a bacterial  
PT infection.  
XX Example 1; SEQ ID NO 5384; 243pp; English.  
XX The invention relates to an isolated nucleic acid derived from  
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridising to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 364 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids is useful for diagnosing pathological conditions  
CC resulting from E. faecium bacterial infection (e.g. urinary tract

CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
CC (infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of Candida albicans -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating Enterococcus faecium infections. The present sequence represents  
CC one of the disclosed E. faecium proteins.  
XX  
SQ Sequence 341 AA;

Query Match 9.1%; Score 234.5; DB 7; Length 341;  
Best Local Similarity 28.5%; Pred. No. 3.2e-06;  
Matches 96; Conservative 44; Mismatches 138; Indels 59; Gaps 16;

QY 190 DQKTVSV-----GDTVKYIT-----YKNVNHGTEKYQYVIKDTMPSASVVDLNEGS 239  
DB 15 EQETISYDLRGKTASYITAPIPFVDSVLENGSAVIKNYKIDT---PTVGLTYVD 70  
QY 240 YEVITDGSNITTLTQSEKATGKNLLENNNTIIPWAATPTGNTONGANDDF 299  
DB 71 QIEVRAGE---TILTKGD-----YIVEVWSNGFVVTILTBEENGVAKVDTLGLRAD--- 119  
QY 300 YGINTITVYTGVLKSGAKPGSADLPENTNATINPNTSND-----DPGOKVTVRDGOI 354  
DB 120 ARG-GDLITYN--LKVSTELADPHNNTAVIEGRNDEFYERGVEPEKVT--GGR 174  
QY 355 TIKKIDGSTKASLOGAIFVLKNA-TGQFLNF-----NDTNVWEGTEANATBYT 402  
DB 175 KPEKYDASSSELKDKARPELWNEDESRVAIFYKGSPLAVYBSGADRIEWATSGQATEFV 234  
QY 403 TGADGIITITLKEGTYYLVEKAPLGNLLDNS---QKVLGDCATDTTSDNLLV--- 456  
DB 235 ADNGGYFEQGLDYGTYQMKETMAPEGYVLPTEGAATFEFIISYG---SYNEEIQIVGE 291  
QY 457 NP---TVENNKTELPSTGGIGCTTPIYIIGAILVIGA 490  
DB 292 NPGPERVPMKRGSLPATGGNGLLAFLIGISLMIGA 328

RESULT 20  
ABB47318  
ID ABB47318 standard; protein; 793 AA.

XX ABB47318;  
AC ABB47318;  
DT 05-FEB-2002 (first entry)  
DE Listeria monocytogenes protein #22.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KW vitamin B12; bacterial infection; disease.  
XX  
OS Listeria monocytogenes.

XX  
XX  
XX WO200177335-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-FR001118.  
XX  
XX 11-APR-2000; 2000FR-00004629.  
XX  
XX (INSP ) INST PASTEUR.  
XX  
XX Buchrieser C, Frangeul L, Couve E, Ruaniock C, Fsihi H, Dehoux P;  
XX Dusurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;  
XX Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;  
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;  
XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
XX Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;  
XX Rose M, Voss H;

DR WPI; 2002-010914/01.  
XX  
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and related  
PT polypeptides.

XX  
PS Claim 6; SEQ ID NO 23; 192pp; French.  
XX  
CC The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 793 AA;

Query Match 9.0%; Score 232; DB 5; Length 793;  
Best Local Similarity 25.1%; Pred. No. 1.3e-05;  
Matches 125; Conservative 66; Mismatches 189; Indels 118; Gaps 23;

QY 45 TYKAYKVF--DAEIDNANVSDNKGASYLIPQ-----GKEAEYK---- 82  
DB 239 TYSADRVFTAGAPISSTNFS-ATSDGFSVALGNLTDVQISYITTTTDDGGSTQDNTAK 297  
QY 83 -ASTDF-----NSLFTTTTNG-----GRYTVTKD--TASANEIATW-----A 117  
DB 298 LAGTDFVTQKSTWTWPASGGGEGGTTGTVLTKEADAKTKATLEGAEFKLVDSKGTVLQ 357  
QY 118 KSISANTPTPVSTVETESNNDGTEVINVSQYGYVYSST-----VNNCAVIMVTSVTPNA 170  
DB 358 ENITTNASGQLSIADLPDFTYQLIETKAPTGYKLDTPVEFTIGENNOAI----TVTKEN 413  
QY 171 TIHEKNTDATWGDGGK--TVQKTVSVGDTVKYITTYKNAVNVHTEKV-----YQVY 222  
DB 414 TLNTGSVELTKLDAATKATLAGATFELQDKEGNTLQDLDKTDENGVLKVTDLVPGSYQFV 473  
QY 223 IKDTMPSASVVDLNEGSYEVITDGSNITTLTQSEKATGKNLLENNNTIIPWAA 282  
DB 474 -ETSAPTGYKLDNSPVSPFV-IAGETDQVVKVT-----KENTLEVG-----SVELTK 518  
QY 283 TMTPTGNTQANDDDFFYKGINITITVTV-----TGVLK--SGAKPGSADLPEN---TNIATI 334  
DB 519 LDSATKATLAGATFELQDKEGNTLQDLDKTDENGVLKVTDLVPGSYQFVTSAPTGYKLD 578  
QY 335 NPNTS-----NDDPGOKVTVRD-----GOITIKKIDGSTKASLOGAIFVLKNATQOFLNF 384  
DB 579 NSPVSPFVWAGETDQVVKVTENTLEVGSEVLTKLDSATKATLAGATFELQDKEGNTLQ- 637  
QY 385 NDTNNVWEGTEANATEYTTGADGIITITGLKEGTYYLVEKAPLGNLLDNSQKVLGDD 444  
DB 638 -----TGLTTDENGVLKVTDLVPGTVQFVETKAPIGELYELDTTPVSPFVAVG 683  
QY 445 ATD---TTNSDNLLVNPT 459  
DB 684 ETDPVTKVTKENTLVPPT 701  
RESULT 21  
ABU32765  
ID ABU32765 standard; protein; 793 AA.

XX	ABU32765;	Matches 125; Conservative 66; Mismatches 189; Indels 118; Gaps 23;
AC	19-JUN-2003 (first entry)	45 TYKAYKVF--DAEIDNANVSDSNKDGASYLIPO-----GKEAEYK---- 82
DT	Protein encoded by Prokaryotic essential gene #18292.	239 TYSADRVFTAGAPISSTNFS-ATSDGFSVALGNLUTDSVQISYTTTDDGKSGTQYDNTAK 297
XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.	83 -ASTDF-----NSLFTTTTNG-----GRTYVTKD--TASANEIATW-----A 117
XX	Listeria monocytogenes.	298 LAGTDFVTWKQTSWTWPASGGGGGGTTSVTLTKEDAKTKATLEGAEFKLVDKSGTVLQ 357
XX	WO200277183-A2.	118 KSISANTPTVSTVTSNNDGTEVINVSQYGYYSST-----VNNGAVIMVTSVTPNA 170
XX	03-OCT-2002.	358 ENITTNASGQLSIADLKFDFTYQLIETKAPTGYKLDTPVEFTIGENNOAI-----TVTKEN 413
XX	21-MAR-2002; 2002WO-US0009107.	171 TIHEKNTDATWGDGGK--TVDQKTVSVGDTVKYTIITYKNVNVHGTBKV-----YQYV 222
XX	21-MAR-2001; 2001US-00815242.	414 TLNTGSELTKLDAATKATKATLAGATFELQDKGNTLQTDLDKTDENGVLKVDLVPGSQFV 473
XX	06-SEP-2001; 2001US-00948993.	223 IKDTMPSASVVDLNEGSVEYITDGSNITLTGSEKATGKYNLLBENNFTIIPWAA 282
XX	25-OCT-2001; 2001US-0342923P.	474 -ETSAPTGYKLDNSPVSEV- IAGETDQVVKVT-----KENTLEVG-----SVELTK 518
XX	08-FEB-2002; 2002US-00072851.	283 TNTPTGNTQNGANDDFYKGINITITVY-----TGVLK--SGAKPGSADLPEN---TNIATI 334
XX	06-MAR-2002; 2002US-0362699P.	519 LDSATKATLAGATFELQDKGNTLQTDLDKTDENGVLKVDLVPGSQFVETSAFTGYKLD 578
XX	(ELIT-) ELITRA PHARM INC.	335 NPNTS-----NDDPGQKVTVRD-----GQITIKKIDGSTKASLOGAIFVLKXNATGQFLNF 384
XX	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;	579 NSPVSFEVAGETDQVVKVTENTLEVGSELTKLDSATKATLAGATFELQDKGNTIQ- 637
XX	Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;	385 NDTNNVWGTEANATEYTTGADGIITITGLKEGYLYLVEKKAPLGYNLLDNSQKVLGDG 444
XX	WPI; 2003-029926/02.	638 -----TGLTDTDENGVLKVDLVPGTGYQFVETKAPIGYELDTTPTVSEFVAG 683
XX	N-PSDB; ACA36635.	445 ATD-----TTNSDNLLVNPT 459
XX	New antisense nucleic acids, useful for identifying proteins or screening	684 ETDPIVKVTENTLVPPT 701
XX	for homologous nucleic acids required for cellular proliferation to	RESULT 22
XX	isolate candidate molecules for rational drug discovery programs.	AAM16315
XX	Claim 25; SEQ ID NO 60689; 1766pp; English.	ID AAM16315 standard; protein; 688 AA.
XX	The invention relates to an isolated nucleic acid comprising any one of	AC AAM16315;
XX	the 6213 antisense sequences given in the specification where expression	DT 12-OCT-2001 (first entry)
XX	of the nucleic acid inhibits proliferation of a cell. Also included are:	XX Peptide #2749 encoded by probe for measuring cervical gene expression.
XX	(1) a vector comprising a promoter operably linked to the nucleic acid	XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX	encoding a polypeptide whose expression is inhibited by the antisense	XX cervical cancer.
XX	nucleic acid; (2) a host cell containing the vector; (3) an isolated	XX Homo sapiens.
XX	polypeptide or its fragment whose expression is inhibited by the	XX WO200157278-A2.
XX	antisense nucleic acid; (4) an antibody capable of specifically binding	XX 09-AUG-2001.
XX	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	XX 30-JAN-2001; 2001WO-US0000670.
XX	proliferation or the activity of a gene in an operon required for	XX 04-FEB-2000; 2000US-0180312P.
XX	proliferation; (7) identifying a compound that influences the activity of	XX 26-MAY-2000; 2000US-0207456P.
XX	the gene product or that has an activity against a biological pathway	XX 30-JUN-2000; 2000US-00608408.
XX	required for proliferation, or that inhibits cellular proliferation; (8)	XX 03-AUG-2000; 2000US-00632366.
XX	identifying a gene required for cellular proliferation or the biological	XX 21-SEP-2000; 2000US-0234687P.
XX	pathway in which a proliferation-required gene or its gene product lies	XX 27-SEP-2000; 2000US-0236359P.
XX	or a gene on which the test compound that inhibits proliferation of an	XX 04-OCT-2000; 2000GB-00024263.
XX	organism acts; (9) manufacturing an antibiotic; (10) profiling a	XX (MOLR-) MOLECULAR DYNAMICS INC.
XX	compound's activity; (11) a culture comprising strains in which the gene	XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX	product is overexpressed or underexpressed; (12) determining the extent	XX WPI; 2001-488901/53.
XX	to which each of the strains is present in a culture or collection of	
XX	strains; or (13) identifying the target of a compound that inhibits the	
XX	proliferation of an organism. The antisense nucleic acids are useful for	
XX	identifying proteins or screening for homologous nucleic acids required	
XX	for cellular proliferation to isolate candidate molecules for rational	
XX	drug discovery programs, or for screening homologous nucleic acids	
XX	required for proliferation in cells other than S. aureus, S. typhimurium,	
XX	K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of	
XX	the target prokaryotic essential genes. Note: The sequence data for this	
XX	patent did not form part of the printed specification, but was obtained	
XX	in electronic format directly from WIPO at	
XX	ftp.wipo.int/pub/published_pct_sequences	
XX	Sequence 793 AA;	
XX	Query Match 9.0%; Score 232; DB 6; Length 793;	
XX	Best Local Similarity 25.1%; Pred. No. 1.3e-05;	





XX AAM28810;  
AC 17-OCT-2001 (first entry)  
DT  
DE Peptide #2847 encoded by probe for measuring placental gene expression.  
XX  
XX Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200157272-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000663.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human placenta.  
XX  
XX Claim 27; SEQ ID NO 29079; 654pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SNP:  
CC see AA131315-AA157546). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for producing a microarray for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from human placenta. The probes are useful for antenatal diagnosis of  
CC human genetic disorders  
XX  
XX Sequence 688 AA;  
Query Match 8.6%; Score 222.5; DB 4; Length 688;  
Best Local Similarity 24.3%; Pred. No. 4.1e-05;  
Matches 115; Conservative 57; Mismatches 221; Indels 81; Gaps 18;  
QY 32 TGTITVODTQKATYKAYKVFDAEIDIANVSDSNKDCASYLIPOGKEAEYKASTDFNSLF 91  
DB 38 TGSEITKVSTGSETTTTSTEGSEITATISITGETTTAS--TEGSETT-TASTEGSETT 93  
QY 92 TTTTNGRTVYVKDTSANEAIA-----TWAKSISANTPVTSTVSTESNNDGTEVINVS 144  
DB 94 SASITGSET-TTASTTSSETTASIMGSETTMASTIGSETTKVST---ASSKMTTVF--- 146  
QY 145 QYGYVYSSVTNNGAVLWTSVTPNATIEHKNTDATWGDGGKTVDQKTVSVGDTVKYTI 204  
DB 147 -----TENSETTASTASTSTTTVSTAGSETTIPASTAGSETTTTSTEGSETTAS 197  
QY 205 TYKNVANYHGTEKYVYVVIDTWPESASVDLNEGSYEVTI-TDQSGNITTLTGSEKATG 263  
DB 198 TEGSETTTASTES-----SETTTATTI-----GSETTTASTEGSETTTTSTEGSETTTA 246  
QY 264 KYNLENNNFTTIPWAATNTPGNTONGANDFFPKGINTVITVTGVLKSGAKPGSA 323  
DB 247 S-----TEGSEIT-TVSTGSETTTTASTEGSETTTASTEGSELTTVSTTG---SETITVSA 298  
QY 324 DLPTNTIATINPNTSNDPG-----QKVTVRDGQITIKKIDGS--TKASLQGAIFVLKNA 377  
DB 299 EGSETTTVTMGSETTTASTAGSETTTVSTAGSETTTASTAGSETTTVSTGSETTTVST 358

QY 378 TGOFLNFNDTNNVW-----GTEANATEYTTGADGIITITLKEGTIVYLVEKAPLGYN 431  
DB 359 TGTETTTTSTEGSEITTTVTAGSETTAV-YTTGSETTTTSTEGSETT----- 404  
QY 432 LLDSNQKVLGSGADTTTNSDNLVNPTVE--NNKGTELPSTGGTGTTTFYIIGA 484  
DB 405 -----TVSTTGTSETTTTASTADLETTTTSTGSGTTTTASTAGSETTTVTYITGS 451  
RESULT 25  
ABB30137  
ID ABB30137 standard; peptide; 688 AA.  
XX  
XX AC ABB30137;  
XX  
XX DT 01-FEB-2002 (first entry)  
XX  
XX DE Peptide #2788 encoded by breast cell single exon nucleic acid probe.  
XX  
XX KW Human; microarray; single exon probe; gene expression; breast; disease;  
KW cancer.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157271-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US000662.  
XX  
XX PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-496933/54.  
XX  
XX New spatially-addressable set of single exon nucleic acid probes, useful  
PT for measuring gene expression in sample derived from human breast,  
PT comprises number of single exon nucleic acid probes.  
XX  
XX Claim 27; SEQ ID NO 13105; 327pp + Sequence Listing; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting the  
CC probes with a collection of detectably labelled nucleic acids derived  
CC from mRNA of human breast, and then measuring the label bound to each  
CC probe of the microarray. The probes are useful for verifying the  
CC expression of regions of genomic DNA predicted to encode proteins. They  
CC are useful for gene discovery, and for determining predisposition and/or  
CC prognosing breast disease. Gene expression analysis is useful for  
CC assessing the toxicity of chemical agents on cells. The microarray of  
CC this invention presents a far greater diversity of probes for measuring  
CC gene expression, with far less bias than expressed sequence tag  
CC microarrays. The method is suitable for rapid production of functional  
CC information from genomic sequence. The present sequence is a peptide  
CC encoded by a single exon nucleic acid probe of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 688 AA;  
Query Match 8.6%; Score 222.5; DB 4; Length 688;



PD	09-AUG-2001.	AC	AAW56138;
XX		XX	
PF	30-JAN-2001; 2001WO-US000668.	DT	05-NOV-2001 (first entry)
XX		XX	
XX	04-FEB-2000; 2000US-0180312P.	DE	Human brain expressed single exon probe encoded protein SEQ ID NO: 28243.
PR	26-MAY-2000; 2000US-0207456P.	XX	
PR	30-JUN-2000; 2000US-00608408.	KW	Human; brain expressed exon; gene expression analysis; probe; microarray;
PR	03-AUG-2000; 2000US-00632366.	KW	Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
PR	21-SEP-2000; 2000US-0234687P.	XX	
PR	27-SEP-2000; 2000US-0236359P.	OS	Homo sapiens.
PR	04-OCT-2000; 2000GB-00024263.	PN	WO200157275-A2.
XX		XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.	PD	09-AUG-2001.
XX		XX	
XX	Penn SG, Hanzel DK, Chen W, Rank DR;	PF	30-JAN-2001; 2001WO-US000667.
XX		XX	
XX	WPI; 2001-488900/53.	XX	
XX		PR	04-FEB-2000; 2000US-0180312P.
XX		PR	26-MAY-2000; 2000US-0207456P.
XX		PR	30-JUN-2000; 2000US-00608408.
XX		PR	03-AUG-2000; 2000US-00632366.
XX		PR	21-SEP-2000; 2000US-0234687P.
XX		PR	27-SEP-2000; 2000US-0236359P.
XX		PR	04-OCT-2000; 2000GB-00024263.
XX		XX	(MOLE-) MOLECULAR DYNAMICS INC.
XX		XX	Penn SG, Hanzel DK, Chen W, Rank DR;
XX		PI	
XX		XX	
XX		XX	WPI; 2001-483446/52.
XX		DR	
XX		XX	
XX		XX	Single exon nucleic acid probes for analyzing gene expression in human
XX		PT	brains.
XX		XX	
XX		XX	
XX		PS	Example 4; SEQ ID NO 28243; 650pp + Sequence Listing; English.
XX		XX	
XX		CC	The present invention provides a number of single exon nucleic acid
XX		CC	probes which are derived from genomic sequences expressed in the human
XX		CC	brain. They can be used to measure gene expression in brain cell samples,
XX		CC	which may enable the diagnosis and improved treatment of nervous system
XX		CC	diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX		CC	epilepsy and cancers. The present sequence is a protein encoded by one of
XX		CC	the probes of the invention
XX		XX	
SQ	Sequence 688 AA;	SQ	Sequence 688 AA;
			Query Match 8.6%; Score 222.5; DB 4; Length 688;
			Best Local Similarity 24.3%; Pred. No. 4.1e-05;
			Matches 115; Conservative 57; Mismatches 221; Indels 81; Gaps 18;
QY	32 TGTITVDDTQKATYKAYKVFDAIDNANVSDNKGASLYLIPQKEAEYKASTDFNSLF 91	QY	32 TGTITVDDTQKATYKAYKVFDAIDNANVSDNKGASLYLIPQKEAEYKASTDFNSLF 91
DB	38 TGSSETTKVSTGSETTTTSGSETTTASTGSETTTASITGSETTTAS 93	DB	38 TGSSETTKVSTGSETTTTSGSETTTASTGSETTTASITGSETTTAS 93
QY	92 TTTTNGGRTYVTKDTSANAEIA-----TWAKSISANTPVSIVTESNNDGTEVINVS 144	QY	92 TTTTNGGRTYVTKDTSANAEIA-----TWAKSISANTPVSIVTESNNDGTEVINVS 144
DB	94 SASITGSET-TTASTTSSETTWSIMSGSETTWASITGSETTKVST---ASSKMTTVF--- 146	DB	94 SASITGSET-TTASTTSSETTWSIMSGSETTWASITGSETTKVST---ASSKMTTVF--- 146
QY	145 QYGYVVSSTVNGAVIMVTSVTPNATHEKNMTDATWGDGCKTVDOQKTVSGDVKYTI 204	QY	145 QYGYVVSSTVNGAVIMVTSVTPNATHEKNMTDATWGDGCKTVDOQKTVSGDVKYTI 204
DB	147 -----TENSETTIASTTASTTASTTASTTASTTASTTASTTASTTASTTASTTASTTAST 197	DB	147 -----TENSETTIASTTASTTASTTASTTASTTASTTASTTASTTASTTASTTASTTAST 197
QY	205 TYKNVNVHGTKEKYQYVVKDTPMSASVVDLNEGSYEVTI-TDGSNITTLTGSEKATG 263	QY	205 TYKNVNVHGTKEKYQYVVKDTPMSASVVDLNEGSYEVTI-TDGSNITTLTGSEKATG 263
DB	198 TEGSETTTASTES-----SETTTATTI-----GSETTTASTEGSETTTASTEGSETTTA 246	DB	198 TEGSETTTASTES-----SETTTATTI-----GSETTTASTEGSETTTASTEGSETTTA 246
QY	264 KYNLLBNNNFTITIIPWAATNPTGNTQNGANDFFYKGINITVTVTVGLKSGAKPGSA 323	QY	264 KYNLLBNNNFTITIIPWAATNPTGNTQNGANDFFYKGINITVTVTVGLKSGAKPGSA 323
DB	247 S-----TEGSEIT-TVSTTSGSETTTASTEGSETTTASTEGSETTTASTTGTG---SETITVSA 298	DB	247 S-----TEGSEIT-TVSTTSGSETTTASTEGSETTTASTEGSETTTASTTGTG---SETITVSA 298
QY	324 DLPENTNIATNPNTSNDPPG-----QKVTVRDQGITIKKIDGS--TKASIQGALFVLKNA 377	QY	324 DLPENTNIATNPNTSNDPPG-----QKVTVRDQGITIKKIDGS--TKASIQGALFVLKNA 377
DB	299 EGSETTTVTVMGSETTTASTAGSETTTVSTAGSETTTASTAGSETTTASTAGSETTTAST 358	DB	299 EGSETTTVTVMGSETTTASTAGSETTTVSTAGSETTTASTAGSETTTASTAGSETTTAST 358
QY	378 TQGFNFNDNNV-----GTENATEYTTGADGIITITGLKEGYLLVEKAPLGYN 431	QY	378 TQGFNFNDNNV-----GTENATEYTTGADGIITITGLKEGYLLVEKAPLGYN 431
DB	359 TGTETITITSGSETTTVTAGSETTAV-YTTGSETTTTSGSETTT----- 404	DB	359 TGTETITITSGSETTTVTAGSETTAV-YTTGSETTTTSGSETTT----- 404
QY	432 LLDNSQKVLGDGATDNTNSDLNVPVE--NNKTELPSTGGIGTITFIIGA 484	QY	432 LLDNSQKVLGDGATDNTNSDLNVPVE--NNKTELPSTGGIGTITFIIGA 484
DB	405 -----TVSTGSETTTASTADLETTTVTSGSETTTASTAGSETTTASTAGSETTTAST 451	DB	405 -----TVSTGSETTTASTADLETTTVTSGSETTTASTAGSETTTASTAGSETTTAST 451
			RESULT 28
			AAW56138
			ID AAW56138 standard; protein; 688 AA.
			XX



CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridises at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosing  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 688 AA;

Query Match 8.6%; Score 222.5; DB 4; Length 688;  
Best Local Similarity 24.3%; Pred. No. 4.1e-05;  
Matches 115; Conservative 57; Mismatches 221; Indels 81; Gaps 18;  
Qy 32 TGTITVQDTQKGYKAYKVFDAEIDNANVSDSNKDGASYLIPOGKEAYEKASTDFNSLF 91  
Db 38 TGSETTKVSTGSETTTTSTEGSEITITASITGSETTAS---TEGETT-TASTEGSETT 93  
Qy 92 TTTTNGGRITYVTKKDTASANEIA-----TWAKSISANTPVSTVTESNNDGTEVINVS 144  
Db 94 SASITGSET-TTASTTSSETTASIMGSETTMASTIGSETTKVST---ASSKMTTVF--- 146  
Qy 145 QYGYVYVSVNNGAVIMVTSVTPNATIEHKNTDATMGDGGKTVDOKTVSVGDTVKYTI 204  
Db 147 -----TENSETTIASTAGSETTTVSTAGSETTIPASTAGSETTTTSTEGSETTAS 197  
Qy 205 TYKNVNVYHGEKYQYQVVKDTPSASVVDLNEGSVEVTI-TDGSQNIITLTQSEKATG 263  
Db 198 TEGSETTTASTES-----SETTATII-----GSETTTASTEGSETTTTSTEGSETTTA 246  
Qy 264 KYNLLENNNFTITIPWAATNPTGNTONGANDDFYKGINITITVYTVGLKSGAKPGSA 323  
Db 247 S-----TEGSEIT-TVSTTGSTTTASTEGSETTTASTEGSELTAVSTTG---SETITVSA 298  
Qy 324 DLPENTNIATINPNTSNDPG---QKVTVRDQGITIKKIDGS--TKASLQGAIFVLKNA 377  
Db 299 EGSETTTVTGSETTTASTAGSETTTVSTAGSETTTASTEGSETTTVSTGSETTTVST 358  
Qy 378 TGQFLNFNDNNVEW-----GTEANATEYTTGADGIITITGLKEGYLVLEKAPLGYN 431  
Db 359 TGTETTTTSTEGSETTTVTGSETTAV-VTTGSETTTTSTEGSEIT----- 404  
Qy 432 LLDNSQVILGDGATDTNSDNLVNPTVE-NNKGTLPSTGGIGTTFYIIGA 484  
Db 405 -----TVSTTGSTTTASTADLETTVTSGSGITTASTAGSETTTVITGS 451

Search completed: July 22, 2004, 10:27:32  
Job time : 60 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:21:56 ; Search time 19 Seconds  
(without alignments)  
2541.481 Million cell updates/sec

Title: US-10-009-254-2  
Perfect score: 2578  
Sequence: 1 MKKMIQSLVSLAFGMVAV.....GALLVIGAGIVLVARRLRS 502

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 90 summaries

Database :

PIR 78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284	11.0	537	2 A35400	surface protein T6
2	265.5	10.3	665	2 F95053	cell wall surface
3	258.5	10.0	614	2 F86719	hypothetical prote
4	258	10.0	533	2 A35259	type 1 fimbrial pr
5	251	9.7	507	2 S52348	hypothetical prote
6	232	9.0	793	2 AH1094	probable peptidogl
7	231.5	9.0	534	2 A32347	fimbrial protein t
8	219	8.5	1612	2 AB1335	probable peptidogl
9	217	8.4	562	2 AE1335	probable peptidogl
10	216	8.4	1806	2 AF1717	probable peptidogl
11	214.5	8.3	681	2 AC1458	surface anchored p
12	212	8.2	2551	2 B98047	hypothetical prote
13	205	8.0	1622	2 AE1717	probable cell surf
14	200.5	7.8	893	2 E95053	cell wall surface
15	200	7.8	1441	2 B86807	hypothetical prote
16	199	7.7	1530	2 AH1396	peptidoglycan anch
17	192.5	7.5	571	2 AI1094	probable peptidogl
18	191	7.4	1711	2 AB1283	peptidoglycan link
19	189	7.3	1519	2 S41525	major ring-forming
20	186.5	7.2	1275	2 T33369	hypothetical prote
21	186	7.2	3705	2 AD0123	probable autotrans
22	185.5	7.2	583	2 A48995	cspB protein - Clo
23	185.5	7.2	1026	2 A48995	patacrystalline su
24	185.5	7.2	1073	2 C87374	S-layer protein Rs
25	184.5	7.2	2468	2 AB3412	hypothetical prote
26	181.5	7.0	1315	2 T28679	fibrinogen-binding
27	181	7.0	3029	2 S76109	hypothetical prote
28	180	7.0	1166	2 T28680	fibrinogen-binding
29	180	7.0	1231	2 T28681	rib protein - Stre

#### RESULT 1

A35400  
Surface protein T6 precursor (strain D471) - Streptococcus pyogenes  
C:Species: Streptococcus pyogenes  
C:Date: 31-Aug-1990 #sequence\_revision 31-Aug-1990 #text\_change 15-Oct-1999  
C:Accession: A35400  
R:Schneewind, O.; Jones, K.F.; Fischetti, V.A.

#### ALIGNMENTS

30	180	7.0	1286	2 S28634	adhesin AIDA-I pre
31	179.5	7.0	2021	2 A97859	190-KDa cell surfa
32	178.5	6.9	4936	2 AH2515	hypothetical prote
33	178	6.9	1041	2 S55862	probable membrane
34	177.5	6.9	1185	2 A42404	collagen adhesin -
35	177.5	6.9	1965	2 S75200	fat protein - Syne
36	177	6.9	1029	2 T30852	outer membrane pro
37	176.5	6.8	1385	2 D89824	hypothetical prote
38	176	6.8	1276	2 B86546	polymorphic outer
39	176	6.8	1276	2 C81591	polymorphic membra
40	175	6.8	1536	2 A43855	high-molecular-wei
41	175	6.8	1999	2 AB2018	hypothetical prote
42	174.5	6.8	1461	2 E90696	hypothetical prote
43	174.5	6.8	1461	2 E90696	hypothetical prote
44	174.5	6.8	5188	2 B85547	probable RTX fami
45	174.5	6.8	5291	2 F90696	hypothetical prote
46	174	6.7	1428	2 AC2224	hypothetical prote
47	173.5	6.7	586	2 AD1458	probable peptidogl
48	173	6.7	359	2 S42787	serine/threonine-r
49	173	6.7	1528	2 D85912	hypothetical prote
50	173	6.7	1571	2 C91068	hypothetical prote
51	172.5	6.7	393	2 G95053	cell wall surface
52	172	6.7	953	2 C89824	hypothetical prote
53	171.5	6.7	1407	2 B72078	polymorphic outer
54	171.5	6.7	1983	2 G86643	hypothetical prote
55	171.5	6.7	2354	2 AB3528	extracellular seri
56	171	6.6	595	2 B48658	flagellin - Bacher
57	171	6.6	1661	2 A83695	hypothetical prote
58	169.5	6.6	1348	2 AH1115	cell surface prote
59	169	6.6	1873	2 T30944	surface protein pr
60	169	6.6	2249	2 A41477	190K surface anti
61	168	6.5	821	2 AD1507	probable secreted
62	168	6.5	1910	2 AF0394	probable adhesin h
63	168	6.5	2044	2 AB1180	probable peptidogl
64	168	6.5	3013	2 AB0480	probable invasin y
65	168	6.5	3083	2 AH2493	hypothetical prote
66	167.5	6.5	414	2 AG1459	probable peptidogl
67	167.5	6.5	826	2 AB1841	hypothetical prote
68	167.5	6.5	939	2 I41197	cell surface prote
69	167.5	6.5	1902	2 B45764	lactocarpin (EC 3.4
70	167	6.5	1141	2 E89824	hypothetical prote
71	167	6.5	2020	2 C48399	ABC-type transport
72	166	6.4	821	2 AB1148	probable secreted
73	166	6.4	940	2 AB1744	internalin protein
74	165.5	6.4	949	2 D90803	Aida-I adhesin-lik
75	165.5	6.4	1005	2 H85611	probable adhesin Z
76	165.5	6.4	1588	2 A86036	probable adhesin Z
77	165.5	6.4	1588	2 H91188	probable adhesin E
78	165.5	6.4	1902	2 S06997	lactocarpin (EC 3.4
79	165	6.4	1032	2 T34433	hypothetical prote
80	165	6.4	1386	2 AC1533	surface protein (L
81	165	6.4	1643	2 D71630	outer membrane pro
82	165	6.4	1655	2 E97835	hypothetical prote
83	165	6.4	1816	2 F83901	hypothetical prote
84	165	6.4	2232	2 T34434	hypothetical prote
85	164.5	6.4	794	2 T36972	probable membrane
86	164.5	6.4	1268	2 B99789	hemagglutinin/hemo
87	164.5	6.4	1270	2 E85649	hypothetical prote
88	164.5	6.4	1962	2 A32634	lactocarpin (EC 3.4
89	164	6.4	961	2 AD0548	putative autotransp
90	164	6.4	1569	2 A65044	hypothetical prote

J. Bacteriol. 172, 3310-3317, 1990  
A; Title: Sequence and structural characteristics of the trypsin-resistant T6 surface protein of *Escherichia coli* O157:H7  
A; Reference number: A35400; MUID: 90264329; PMID: 2188957  
A; Accession: A35400  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-537 <SCH>  
A; Cross-references: GB:M32978; NID:G153842; PIDN:AAA27019.1; PID:G153843  
C; Keywords: transmembrane protein

Query Match 11.0%; Score 284; DB 2; Length 537;  
Best Local Similarity 22.2%; Pred. No. 3.7e-08;  
Matches 135; Conservative 89; Mismatches 195; Indels 188; Gaps 28;

[illegible]

RESULT 2  
F95053  
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain T)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: F95053  
R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Lofcus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95053

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-865 <KUR>  
A:Cross-references: GB:AS005672; PTDN:AAK74623.1; PTD:GI4971934; CSPDB:GN00164; TIGR:SP48  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0463

Query Match	10.3%	Score 265.5;	DB 2;	Length 665;
Best Local Similarity	24.4%;	Pred. No. 4.8e-07;		
Matches 164:	Conservative	63;	Mismatches 227;	
			Indels 219;	Gaps 33;

Qy	9	LLVASLAFGWAVSPVPIAFAAETGTTTVODTQKGATYKAYKVDFDAIDIDNANVSDSNKDG	68
Db	15	LLTASSLFSAA-----VFAAGTTTTSVTHKLLATDGMDKTANLETTGNYA-GNKVG	67
Qy	69	ASYLIPQKGAEYKASDFNSLFTTTN-----GGRTYVTKDTPASANEIATWAKSISAN	123
Db	68	---VLP-----ANAKEIAGVWFVNTNNEIIDENGQTLGVNIDP-----QTFKLSGAMP	114
Qy	124	TPVSVTVTESNNDCGTEVINV--SOYGYVYVSS-----TVNNGAVIMVTSVTP-----	168
Db	115	ATAMKLTAEAGAKFANTANLPAAYKYIEHLSLTSYVGEDGATLTKGKAVPIETIELPLND	174
Qy	169	--NATIEHKNTDA---TWGDDGGKT-----VDOKT---YSGVDTKVYTIYK--NAVNY	212
Db	175	VVDARVYPKNTEAKPKIDKDFKGANPDTPRVDKDPVNHQGVGDVVEYIVTKIIPALANY	234
Qy	213	-----HGTEKVY-----QYVIKDT-----MPSASVVDLNEG--	238
Db	235	ATANWSDRMTBGLAFNKGTVKVTVDVDALEAGDYALTEVATGFDLKLTDAGLAKVNDQNA	294
Qy	239	-----SYEVTIITDGS-----GNITLITGQSEKATG--KYNLEENNNFTIIPWA-A	282
Db	295	EKTVKIYTSATLNDKAIIVEPESNDVTFNYGNPDHGNTPKPNKPNENGDUJLTKTWVDA	354
Qy	283	TNTPTG-----NTQGANDDFFYKGINTI-----TVTYTGVLKS-----	317
Db	355	TGAPIPAGABATFDLVNAQTG-----KVQVTVTLTKNTVTVNGLDKNTYKFVVERS	407
Qy	318	AKPSADLIPENTNIATI-----NPNTSNDPPGQKVTVRDCQIITIKKIDGSTKASLQCAI	371
Db	408	IKGYSADYQEITTAGIEIAVRNWKNDENPKLPDTEPKVTVYTKKFVKVNDKDNR--LAGAE	465
Qy	372	FVLXNA--TGQFL-----	382
Db	466	FVIANADNAGQYLARKADKVSQEBEQKLVVTTKDALDRAVAAYNALTAQQQTQOEKEKVDK	525
Qy	383	---NPN-----DTNVEWGTB--ANATEYTTGADGIIITITGLKEGTYYLVEKKAPLGYN	431
Db	526	AQAAYNAAVIAANNAFENWADKDNENVVVKLVSDAQGRFEITGLLAGTYVLYEETKQPAGYA	585
Qy	432	LLDNSOKV-----ILGDGATDTTSDNLLVNPTVNNKGTLPSTGGTGTTFYFIIG	483
Db	586	LLTSRQKPEVTVATSYSGTQGIETYAGSGK--DDATKVVNKKIITIPQTGGGTIIFAVAG	643
Qy	484	AILVIGAGIILVA	496
Db	644	AAIM--GIAVYA	653

RESULT 3  
F86719  
hypothetical protein yhgE [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C.Species: Lactococcus lactis subsp. lactis  
C.Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C.Accession: F86719  
R.Polotnin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A.Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp  
A.Reference number: A86625; MUID:21235186; PMID:11337471  
A.Accession: F86719  
A.Status: preliminary



A:Molecule type: DNA  
A:Residues: 1-614 <STO>  
A:Cross-references: GB:AE005176; PID:g12723675; PIDN:AAK04856.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: yhgE

Query Match 10.0%; Score 258.5; DB 2; Length 614;  
Best Local Similarity 24.5%; Pred. No. 1e-06;  
Matches 145; Conservative 57; Mismatches 184; Indels 207; Gaps 31;

```
QY 51 VFDAEIDN-----ANVSDSNKDGASLYLPQGEAEYKASTDFNSLFTTTTNGGRTYVT 103
DB VDSGLANNQTCGGAGTSGTKNDGSGQQTLEGGSTPSTWANTFSAKRVGT-GVPTGVT 133

QY 104 K---KDTASANIATWAKSISANT-----TPVSTVTSNNDGTEVINVSQ----- 145
DB DFSFTDTTSVAPVTTDASGLADFTGLTDGYLYPHQVTVTVNGITTVGDFIVQVSHEDSQAG 193

QY 146 ----YGYVVSSTVNGAVIMVTSVT-----PNAITHE-----KNTD-- 178
DB 194 IVNVYPKLDMSSSAGLG-----TSATTNADDNFGQTFNPQIANPNATGNSDQTLTNTDNN 248

QY 179 -----ATWGDGGGKTVDQKTSYSGDVTYKVTITYKNV-----NVHGTPEKYV-QYV 222
DB 249 AGNENLANGTWNGSDNQ-NTTAAAGNTVNNV---NTVFSSTQTNNGTGTGTGYII 304

QY 223 IKDTMP-----SASVVDLNEGS-----YEVTTIDGSGNI--TTLTQG 257
DB 305 VTDQLPNLNVSSSTVTVSVIVNVINGSGTKVGLTPTTDYIT-NDGNGKIVVTLTTAG 363

QY 258 SKATGKYNLLENNTFTIIPWATNPTGNTQNGANDDPFYKGINITVYTVGLKSG 317
DB 364 QQHAASL--LGSADGALNIIP-STVKSAGSATDSAT-----TTITNAY----- 405

QY 318 AKPGSADLPENTNIATINPNTSDPGQKTVRDGQITIKKIDGSKASLOGAIFVLKNA 377
DB 406 ----GADLSTTAV-----KSTLNVGGLMKTKTDASTNAALAGATFTVWRA 447

QY 378 TQO-----PLNFND---TNNVEWGTANATE-----YTTGAD 406
DB 448 DNKEDAQDFVEANAAYFNNSAGGIVTNLTSSKAFVVTGDTSGNANTSATAPVTTTGKD 507

QY 407 GIITITGLK-----EGT-----YVLEKAPAGYNL-----LDNSQKVLGSGADTT 448
DB 508 GIATENGLNLVDNNTDGSNTTNYLVVEAAPTGYQLPSVTTAANLTGAVTASTAPAAATDT 567

QY 449 TNSDNLVNPTVNNKGTLPSTGGTGTTFYIIGAILVIG--AGIVLVARER 499
DB 568 -----TITNNKFPALPTGGQG-----LAGIIATVSGVIAFAIKR 604
```

RESULT 4  
A35259  
Type 1 fimbrial protein precursor - Actinomyces viscosus  
C:Species: Actinomyces viscosus  
C>Date: 10-Aug-1990 #sequence\_revision 10-Aug-1990 #text\_change 22-Oct-1999  
C:Accession: A35259  
R:Yeung, M.K.; Cisar, J.O.  
J. Bacteriol. 172, 2462-2468, 1990  
A:Title: Sequence homology between the subunits of two immunologically and functionally  
A:Reference number: A35259; MUID:90236904; PMID:1970561  
C:Accession: A35259  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-533 <YEU>  
A:Cross-references: GB:M32067; NID:g141849; PIDN:AAA62572.1; PID:g141850  
C:Keywords: transmembrane protein

Query Match 10.0%; Score 258; DB 2; Length 533;  
Best Local Similarity 26.0%; Pred. No. 9.1e-07;  
Matches 146; Conservative 67; Mismatches 223; Indels 126; Gaps 30;

```
QY 13 SLAFGMVSPVTPPIAFPAETGTITVODTKGATYKAYKVFDABEIDNANVSDSNKDGASYL 72
DB 17 TLAAALVAP-TGAAPADPNNGSTI-DPDAATLTITVHKCEQTDNTNGVKEGTGNE-D----- 69

QY 73 IPOGKEAEYKASTDFNSLFT-----TTTNGGRTVV-TKDDTASANEI-ATWAKSISANT 124
DB 70 -PQ---AECKPVSDEFTITKLVNVDLITYDQWKLADLKGDDVVKAGALKSTIVQKITGA 125

QY 125 TPVSTVTSNNDGTEVINVSQYGYVVSSTVNGAVI-----MVT--SVTP-----N 169
DB 126 NGLASFDTAQ---TEV-----GAYLVSETRTPDKVIPAEDEFVVTLPMTNPQDTAKMNYN 176

QY 170 ATIEHKNTDWTGDDGGKTV-DQKTSYSGDVTYKVTITYK-NANVNHGTPEKYVQYVVKDTM 227
DB 177 VHVYPKNTLS-----GVDKQVTDKPAQSGRDIITITITSIPKVDYFGGARIKRVEVDRL 232

QY 228 -----PSASVVDLNEGSYEVTITDGSNITLTQSGSEKATGKYNLLENNNFIT 277
DB 233 DKRIKKEALTVPVKIV---QNEVTLAE-TTDYTLIT-----AEGK-----DHNWATIQ 277

QY 278 IPWAATNPTGNTQNGANDDPFYKGINT---ITVYTVGLVKSAGKPGSADLPENTNIATI 334
DB 278 LTEGRRKASEARYNGNETKLVTLNAKFDAAVNLEGLDLSNTAGLIPNDSP-NFTWDPN 336

QY 335 NPNTSNDDPG---QKVTVRDQGITIKKI---DGSTKASLOGAIFVL-----KNATGQFLNF 384
DB 337 NPGTTDIPGTPPTPVLSKYKGVLTKTGTDLADTKYNGAQFQVYECTKTASGATLUD 396

QY 385 ND---TNNVEMGTANATEYTTGADGIIITGLK-----EGTVYLVKPKAP 427
DB 397 SDPSTQTVDPITIGGKFTTTAGQGTVEINLVANDYNGAKKOOLDDEDYVCLVETKAP 456

QY 428 LGYN-----LDNSQKVLGSGADTNTSDNLVNPTVNNKGTLPSTGGTGT 477
DB 457 EGYNLQADPLPFRVLAKEAK---KAATEVTVD-----IPKNAGFRLPLTGTANGVI 505

QY 478 IFYIIGAILVIGAGIVLVARER 499
DB 506 FLTTAGALLVAGGAVVAYANKR 527
```

## RESULT 5

S52348  
Hypothetical protein 2 - Lactobacillus leichmannii  
C:Species: Lactobacillus leichmannii  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 05-Nov-1999  
C:Accession: S52348  
R:Schenk-Groeninger, R.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S52347  
A:Accession: S52348  
A:Molecule type: DNA  
A:Residues: 1-507 <SCH>  
A:Cross-references: EMBL:X81869; NID:g666067; PIDN:CAA57459.1; PID:g666069  
A:Experimental source: DSM 20076

Query Match 9.7%; Score 251; DB 2; Length 507;  
Best Local Similarity 24.7%; Pred. No. 2e-06;  
Matches 139; Conservative 78; Mismatches 204; Indels 142; Gaps 27;

```
QY 1 MKKK-----MIQSLVLASLAFGMVSPVTPPIAFPAETGT---ITV----- 37
DB 1 MKSKLSLITGLIMILPLLLLSFATPKQVS-----AADTNNNSVKVTLHKRVFDSQAQEA 53

QY 38 -QDTOK-----GATYKAYKVFDABEIDNANVSDSNKDGASLYLPQGEAEYKAST 85
DB 54 KQNTGEIWNDFGGTGLNGVTFKAYNVTDHYLSLRKSGDSAQDAVTAIOSDAKDSNLPYSY 113

QY 86 DFNSLFTTTTNGGRTYVTK--DTASANEIATWAKSISANTTPVSTVTSNNDGTEVIN- 142
DB 114 AGSATAATETT-----ATSKGEDGIAAFDNLNLKSDGNYQYLVFVETDSPTDVTQQAAP 167

QY 143 -VSQYGYVVSSTVNGAVIMVTSVTPNATHEKNTDWTGDDGGKTVDQK-----TY 194
```



QY	436	SQKVIL--GDGATDTTNSDNLVNPVTVENNKGTLPSTGGIGTTTFYIIG-AILVIGAG	491
Db	464	ASRAITVEPCAGVTQQWIDN--VKQSPG-----LPLTGANGMLTLTASGAALLMIAG	516
QY	492	IVLVARRRLR	501
Db	517	SVLVARYRER	526
RESULT 8			
AB1347			
probable peptidoglycan bound protein (LPXTG motif) lmo2178 [imported] - Listeria monocytogenes			
C:Species: Listeria monocytogenes			
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001			
C:Accession: AB1347			
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.			
Science 294, 849-852, 2001			
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.			
A:Title: Comparative genomics of Listeria species.			
A:Reference number: AB1077; MUID:21537279; PMID:11679669			
A:Accession: AB1347			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1612 <GLA>			
A:Cross-references: GB:NC_003210; PIDN:CAD00256.1; PID:gl6411648; GSPDB:GN00177			
A:Experimental source: strain EGD-e			
C:Genetics:			
A:Gene: lmo2178			
Query Match 8.5%; Score 219; DB 2; Length 1612;			
Best Local Similarity 24.2%; Pred. No. 0.00049;			
Matches 116; Conservative 47; Mismatches 170; Indels 146; Gaps 23;			
QY	43	GATYKAYKVFDAID---NANVSDSNKDGASY--LIPOG-----KEAEYK--ASTD----	86
Db	719	GKGYGSYNTTGNIDWIVSNAMAKYNDLIFDDAIPTLGTYVEGSLQYRNVASTSEMMN	778
QY	87	----FNSLFTTTTNGRTVYTKDKDTASAN---EIAFWAKS---ISANTTPVSTVTSNND	136
Db	779	LYIPLNSVGTVAKTGDKNYPKVDYTGKHLHLEFANLENSRFVFKYSTKP-----	828
QY	137	GTEVINVSQYGYVVSSTV---NNG-----AVIMVTSVTPNATHEKNTDATWG	182
Db	829	-----NENWYFYSVQNTAKVSDNGVGEKSYQAVASKLFNAWTKTATIDP-----	875
QY	183	DGGKTVDQKTVSGDVTKYTYKNAVNYHTEKYQYVVKDTPSASVVDLNEGSEV	242
Db	876	-----SFDKNVNIWTLN---NISADRPINNPTITDTMKT-----GTTGA	912
QY	243	TITDGSNLTTLTGSEKATGKYNLLENNNFITIPMAATNPTGNTQNGANDDFYKKG	302
Db	913	QVKSQFVNETGEDIDSKYDIITYDNNFTIQPK-----DYKA	953
QY	303	INTITYTGV-LKSG-----AKPGSADLPENNTIATINPNTSNDDPGOKV-----T	348
Db	954	TAPIKYTYSTISLMSGLVNTATTASPDYGSLP-----MTYKSRTTISPAFTIGSGST	1008
QY	349	VRDQGITIKKIDGSTKA-SLQGAIFVLKNATGQFLNFNDNNVWGTEANATEYTTGADG	407
Db	1009	ATIGSLEITKVDKONSKKLTGAKFQLYTLEGD-----KAGQEA-----ITDSG	1053
QY	408	IITITGLKEGTYVLEKKAPLGNLLDNSQKVLGDGATDTTNSDNLVNPVTVENNKGT	466
Db	1054	KIVMDGLQSGKYLVEATEAPTGTISDEYK-----DGKEITVTAD-IATSVTIENTEQT	1106
RESULT 9			
AE1335			
probable peptidoglycan bound protein (LPXTG motif) lmo2085 [imported] - Listeria monocytogenes			
C:Species: Listeria innocua			
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001			
C:Accession: AF1717			
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.			

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001			
C:Accession: AE1335			
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.			
Science 294, 849-852, 2001			
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.			
A:Title: Comparative genomics of Listeria species.			
A:Reference number: AB1077; MUID:21537279; PMID:11679669			
A:Accession: AE1335			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-562 <GLA>			
A:Cross-references: GB:NC_003210; PIDN:CAD00163.1; PID:gl6411555; GSPDB:GN00177			
A:Experimental source: strain EGD-e			
C:Genetics:			
A:Gene: lmo2085			
Query Match 8.4%; Score 217; DB 2; Length 562;			
Best Local Similarity 23.2%; Pred. No. 0.00016;			
Matches 131; Conservative 55; Mismatches 197; Indels 182; Gaps 27;			
QY	32	TGTTVQDTQKATYKAYKVPDAEIDNANVSDSNKDGASYL-----IPQKEAEYKAST	85
Db	80	TGNIIEWQNGNGYNYPAF-----SGNFIELNSDGGIPVYQDIRTIP-GSNLTWKFS-	130
QY	86	DNLSLFTTTTNGRTVYTKDKDTA-----SANIA-----TWAKSISANTT-----	125
Db	131	-----HRGRTGV---DTADLLIGSPESQTEVSRVSNGETWG-SPEGNYTVPAGOT	176
QY	126	-----PVSTVTESNNDGTEVINVSQYGYVVSSTVN-NGAVIMVTSVTPNATHEKNT	177
Db	177	ITRLTFNPISTANGSLTSGNFLDDVQLY-----INVNGA-----KIG	213
QY	178	DATWGDGGGKTVQDKTYSVGDVTKYTI-----TYKNVNVHGTKEYQYVVKDTPSASV	232
Db	214	DVVWYDFDNGDGIQDSEEPAPGKVDLLTKDGTFKESAT---TNNIGSYLFTDVLDP----	266
QY	233	VDLNEGSEYEV--TITDGSNITTLTGSEKA-----TGKYNLLEEN---NNFTITIPW	280
Db	267	-----GDQVQFESLPNNDPIFSKANQNDKSLNSKPDKTGIASVNVNPLKSENFDI----	317
QY	281	AATNPTGNTQNG-----ANDDFYKGINITIITYTGV-LKSGAKPG	321
Db	318	-----DAGITTNKGVEIQKLSGDKALSGAVVAIKNSQSEVAKITTNQNGTGAELPGP	372
QY	322	SADLPENTNIATINPNTS-----NDDPGOKVTRD-----GOITIKKIDGSTKASLOG	369
Db	373	NYTATEVTPAPLGQYQKNTTPKFTITYGTPNPKLTFQNAEKTSITIPKQDEANKKGLAN	432
QY	370	ALFVLKNATGQFLNFDNNTNNVWGTEANATEVTTGADGIITITGLKEGYTYLVEKKAPUG	429
Db	433	AVFDVKSIDGTTLK-----KVTNKSQYALAEALNPQGYTIVETVATPPG	476
QY	430	YNLLDNSQKVLGDGATDTTN---SDNLLVNP-----TV-----ENNKGTLPSTG	472
Db	477	YEKSANEIRVTIPFPNQKNTIITFSDNKIMVPLKPTKPTKGVTVVKVSGETTKITALPQTG	536
QY	473	GIGTTFIYIIGAILVIGAGIVLVAR	497
Db	537	DSSSSSTIFTGLLIIVVASGLFVYRR	561
RESULT 10			
AF1717			
probable peptidoglycan bound protein (LPXTG motif) lin282 [imported] - Listeria innocua			
C:Species: Listeria innocua			
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001			
C:Accession: AF1717			
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.			

Science 294, 849-852, 2001  
A;Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A;Title: Comparative genomics of *Listeria* species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AF1717  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1806 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC97510.1; PID:g16414794; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin2282

Query Match 8.4%; Score 216; DB 2; Length 1806;  
Best Local Similarity 23.8%; Pred. No. 0.00083;  
Matches 114; Conservative 50; Mismatches 168; Indels 148; Gaps 19;  
  
QY 43 GATYKAYKVFDAE-----DNANVSDSNKDGASYLIPQKAEAYKASTD--- 86  
Db 721 GGYKAYNSTGNIWDIVSANAMAKSYDNLIFDDTIPTGLTYVEGSLQYRNVESSTEMMS 780  
  
QY 87 -----FNSLFTTTNGGRTYTKDTSANEIATWAKSISANTTPVSTVTSNNDDGTEVI- 141  
Db 781 LQIPLNSTGLAKAGDKNYPTKVDI-TGNKI-----HLEFANLIDNTRVFI 824  
  
QY 142 -----NVSOYGYVYSSTVNGAVIMVTVPNATHEKNTDATWGDGGKTVDOKTYIS 195  
Db 825 KYSTKPDENWFFKYIVN-----IAEVDNGT-----DKHTYS 857  
  
QY 196 -----VGDVTKYITTYKNAVNYHGT-----EKVQYQVKIDTMPASVVDLNE 237  
Db 858 YQAVASVFNAMTKTASIDSTYNNKVNMTVLNISKDRPINNPTTIDTM-----KN 909  
  
QY 238 GSVEVTTDGSNITLTQSEKATGKYNLENNFTIIPWAAATVPTGNTQNGANDD 297  
Db 910 GTTGAQVVKSSFRVNETGGDIDSKYDITFTDKQFIQFK----- 951  
  
QY 298 FFYKGINITVTVTG-VLKSQ-----AKPGSADLPENTNIATINPTNSDDPGOKVT 348  
Db 952 -DYKATDPIKVTYSTVLSMGLSNTATTATSPDYGSLPMSYKSRNTNISPAFTIGSGGT 1010  
  
QY 349 VRDQGITTKID--GSTKASLQGAIFVLKNATGFLNFDNTNNVWGTETANATEYTTGAD 406  
Db 1011 ATIGSLITKVDKDKNTK-KLTGAKFQLYTPEGD-----KAGQEA-----TTDSE 1054  
  
QY 407 GIITITGLKEGTYVLVEKKAPLGNLDNSQKVLGDGATDTTNSDNLVNPVTENKNGT 466  
Db 1055 GKIVMDGLQSGKYLVEAAPTGYTISDEYK-----DGKEITITAD-VATSVTIENTEQ 1108

RESULT 11  
AC1458  
surface anchored protein (LPXTG motif) [imported] - *Listeria innocua* (strain Clip11262)  
C;Species: *Listeria innocua*  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AC1458  
R;Glaser, P.; Frangul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A;Title: Comparative genomics of *Listeria* species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AC1458  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-681 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95435.1; PID:g16414621; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin0202

Query Match 8.3%; Score 214.5; DB 2; Length 681;  
Best Local Similarity 22.0%; Pred. No. 0.00028;  
Matches 110; Conservative 56; Mismatches 194; Indels 141; Gaps 16;  
  
QY 17 GMAVSPVTPIAFAETGTTTODTQKATY-KAYKVFDAE-----IDNANVSDSNK 66  
Db 12 GLAFLSVLIISTVPTTIVKAATSYGSEFLNTVELLDKGTGTPKDFGYDNMNVH---- 67  
  
QY 67 DGASYLIPQKAEAYKASTDF-----NSLFTTTNGGRTYTKDTSANEIATW 116  
Db 68 --YTWSIPNSTNVKAGDSMDPALPSQLALATDLAENVKDSKQTVGTATVTKRATNOV--- 122  
  
QY 117 AKSISANTTPVSTVTSNNND-----GTEVINVSQYGYVYSSTVSN-NG 158  
Db 123 -----TVFSDYVEKHSIDKIGELDFWTFNQKVTGNEKINLE---FPIENSTINVDV 172  
  
QY 159 AVIMVTVPNATHEKNTDATWGGGG-----KTVDOKTYISVGDIVKTYI 204  
Db 173 NVGEKTPVSPTEITFYK-----WVDANNPSLIHWVVRVNYAOKNIPNAVFTDIIGAKQTL 228  
  
QY 205 TVKNAVNVHGTKEKVQYVVIKDTMPSASVVDLNEGSYEVTITDGSNITLTQSEKATGK 264  
Db 229 NFDSTKAFHGTVSTDRVFTAGTPISNTNFSKTSDFSVTL-----GNLTDVQIS 278  
  
QY 265 YNLLBENNFTITIPWAAATNPTGNTQNGANDDFFYKINGINTITVTVTVLKSAGKPGSAD 324  
Db 279 -----YTTATDGGKSTQYDNTAKLSGDFVTKQSTWTWPASGG---GGGADGTT- 325  
  
QY 325 LPENTNIATINPTNSDDPGOKVTVRVDGQITIKKIDGSTKASLQGAIFVLKNATGFLN 384  
Db 326 -----GSVTLTKQDAKTKATGAEKFLVNDANGTITLQE 358  
  
QY 385 NDTNNVWGTETANATEYTTGADGIIITGLKEGTYVLVEKKAPLGNLDNSQKVLGDG 444  
Db 359 N-----ITDASGLNVSNLKATVQLIETKAPTGYKLDSTPVEFTIGEN 403  
  
QY 445 ATD-TTNSDNLVNPVTENNK 464  
Db 404 NKEITVTKENTLDTGAVELTK 424

RESULT 12  
B98047  
hypothetical protein sprl403 [imported] - *Streptococcus pneumoniae* (strain R6)  
C;Species: *Streptococcus pneumoniae*  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C;Accession: B98047  
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; Es  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lerkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.B.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: B98047  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2551 <KUR>  
A;Cross-references: GB:AE007317; PIDN:AAL00207.1; PID:g15459054; GSPDB:GN00174  
C;Genetics:  
A;Gene: sprl403

Query Match 8.2%; Score 212; DB 2; Length 2551;  
Best Local Similarity 23.3%; Pred. No. 0.0021;  
Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;  
  
QY 34 TTTVQDTQKATYKAYKVFDAEIDN-----ANVSDSNKDGASYLI 73  
Db 1891 TFTVQDKGKKGKGRAPKIKVEDITSPIRRDAAATPTRNGIRVTVVDDVNDNGVY-- 1948  
  
QY 74 PQKAEAYKASTDFNSL-----FTTTNGGRTYTKDTSANEIATWAKSISANT 124

Db 1949 DSGVDKVLNSKDIYNGIDGRDGSAPTITTKDNGDGTHTITVQNPDSBSTTVVKDGDGK 2008  
QY 125 TPVSTVTESNNDGTEVINVSQYGYVSVSTVNGAVIMVTSVTPNATTHEKN-----TDA 179  
Db 2009 TANIITTE-NPDGSHITITVNPDPGSKTETVKNKGK-----DGKTPKEVETDNDNDGHTHTVKV 2063  
QY 180 TWGDCG-----GKTVDQKYSVGDVTKYITITYKANVYHGTQKVVYVLKD--- 225  
Db 2064 TDGNGVNTNAIKDGKGAATATTENPDG-SHTVTITNP--DGTKN--EFVYKNGRD 2117  
QY 226 ----TSPASVVDLNEGSVEVITDGSNITTLTQSEKATKYNLLLENNTTITPWA 281  
Db 2118 GVDGRTPASVRDNGSGSHITVITNPEG-VTETTVRDKSPKVTIDB----- 2165  
QY 282 ATNTPGTQNGANDFFYKGINTTITVTVGLKSGAPGSADLPENTNATINPTNSD 341  
Db 2166 -----QNGTHKISVLNGDGTET---IHKDGSFVATVRDQDGTIVRVENGNG 2213  
QY 342 DPGQKVTVRDGO-----ITIKKIDGSTKASLOGAIFVLKNATGQFLNFD 386  
Db 2214 TVSE--TVRDKGSPAKVVDNGDGTHTITVNSDGIITTT-----TVRDKGPEKLEVID 2266  
QY 387 TNNVWEGTEANATEYTTGADGIITITGLKEGTYVLVEKKAPGLYNLLNSQ----- 437  
Db 2267 NN-----DGSHTIKVTGADGKGTITIPDG-----KSPKA-NIVDNGDGTHTLTIVD 2312  
QY 438 -----KVILGSGATDTTNSDNLNPNVTV---ENKNGTEL 468  
Db 2313 SDGREYKSIKDG-KDGKDS-----VSPTVTVKNNNDGTHV 2347  
  
RESULT 13  
AE1717  
probable cell surface protein (LPXTG motif) [imported] - Listeria innocua (strain Clp11  
C:Species: Listeria innocua  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AE1717  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,  
A:Title: Comparative Genomics of *Listeria species*.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AE1717  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1622 <GLA>  
A:Cross-references: GB:AL592022; PIDN:CMC97509.1; PID:gl6414793; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin2281  
  
Query Match 8.0%; Score 205; DB 2; Length 1622;  
Best Local Similarity 22.4%; Pred. No. 0.0028;  
Matches 134; Conservative 65; Mismatches 207; Indels 192; Gaps 27;  
  
QY 18 MAVSPVTPIAFA---AETGTTV-----QDTQK---GATYKAYKVFDAIDNANVSDGNK 66  
Db 991 VAVSPVFTVGTSGGIATIGTIKITKVDDEDDTTKLEGAKFQLY-TLDGKSEQEIT-TNS 1048  
QY 67 DGASYL--IPOGK-----EAAYK-----ASTDFNSLFTTTNG---GRTY 101  
Db 1049 EGEILLDGIQSGKYKLIVETKAPEGYNISDEYKEGKEITVSSGEBLLTIKNAMKGNVI 1108  
QY 102 VTKDQTASANEI-----ATWAKSISANTYPTVSTVESNN----- 135  
Db 1109 LTKDQSASHEVLADAEFELQNATGSKLEKLTDAAGNIEITDLAPGDYKLIETKAPTGY 1168  
QY 136 --DGTETV-----INVSQYGYVSVSTVNGAVIM-----VTSVTPNATTHEKNVTDATWG 182  
Db 1169 QLDATPVNFTIDFNQSEAAKVSKTNTAKTGTVNLTKDQSAWNAELADTFELRNEDETLV 1228

QY 183 DGGGKTVDQKTVSUCD-----TVKYTIITYKNA-----VNYHGTE 216  
Db 1229 RENLVTDDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFITIDFNQSEVEKVSKTNTA 1288  
QY 217 KVVQYVI--KDTMPSASVVDL-----NEGS---YEVTTIDGSGNITTLTQSEKATGKN 266  
Db 1289 KGTGVVLTKDQSATNAELADATFELRNESDILVRENLTDDNGEISV-----ADLAPGDYK 1344  
QY 267 LLEENNNFTITIPWAAATNPTPTGNTQNGANDFFYKGINTTITVTVGLKSGAPGSADLP 326  
Db 1345 LIE-----TKAPAGYQLDAPVNF----- 1363  
QY 327 ENTNATINPNATNSDDPGQKV-TVRDGQITIKKIDGSTKASLOGAIFVLKNATGQFLNPN 385  
Db 1364 -----TIDFNQSEAAKVSKTNTAKTGTVVLTKDQSAWNAELADTFELRNESDILVREN 1417  
QY 386 DTNNVWEGTEANATEYTTGADGIITITGLKEGTYVLVEKKAPGLYNL-----LDNSQ 437  
Db 1418 -----LVTDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFITIDFNQ 1462  
QY 438 KVILGSGATDTTNSDNLNPN-PTVENNK-GTELPSTGGIGTTFYIIGAILVIGAGIV 493  
Db 1463 SEAAKVTNKKKIGITILVNFIDVDGQNLNDKEVHTGNVGBE--YSVKAKEIVGHTLV 1518  
  
RESULT 14  
E95053  
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain TIC  
C:Species: Streptococcus pneumoniae  
C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: E95053  
R:Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide  
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: E95053  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-893 <KUP>  
A:Cross-references: GB:AB005672; PIDN:AAK74622.1; PID:gl4971932; GSPDB:GN00164; TIGR:SP4:  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: Sp0462  
  
Query Match 7.8%; Score 200.5; DB 2; Length 893;  
Best Local Similarity 20.6%; Pred. No. 0.0022;  
Matches 107; Conservative 69; Mismatches 164; Indels 179; Gaps 24;  
  
QY 69 ASYLIPQKKEAYRYKASTDFNSLFTTTTNGGRTYVTKDQTASANEIATWAKSI-----SAN 123  
Db 463 AAFPVKPEKYSKKAAG--YAVIGDPINGGIWL-----NWRESILAYPFNSN 508  
QY 124 TTPVSTVTESNNDGTEVINVSQYGYVSVSTVNGAVIMVTSVTPNATTHEKNVTDATWG 183  
Db 509 TAKIT---NHGDPTP-----WYNGNIAPDGDYDVTVGIGING---DPGTDEATAT 553  
QY 184 GGGKTVDQKT---YSVGDPTVK-----VTITYKANVYHGTQKVVYVLKIDTSPAS 231  
Db 554 SFMQSISSKPENYTNVDTTTLKLEQLNRYFHTIVTEKKSIENTG-----ITDPMGE 604  
QY 232 VVDLNEGS-----YEVTTIDGSG-----GNITTLTQSEK---AT 262  
Db 605 LIDLQGLDGRFPDADYLTLTANDGSRLENGQAVGPGQNDGGLLKNKAKVLYDTTEKIRVT 664  
QY 263 GKYNLLENNFTITIPWAAATNPTPTGNTQNGANDFF-----FYKGINTTITVTVGLKSGA 318  
Db 665 GLYLGTDEKVTLTYNV-----RLNDEFVSNKFF----- 692  
QY 319 KPGSADLPENTNATINP-----NTSNDDPQKQVTVRD-----GQITTKK 358

```
Db 693 -----DINGRTTLHPKVEQNTVRDPIPK--IRDVRKYPEITISKEKLGIDIEFIK 742
QY 359 IDGSTASLOGAIFVLKNAQTQGFLENFNNTNNVWEGTEANAT--EYTTGADGIITITGLKE 416
Db 743 VKNNDKKPLRGAVFSLQK-----OHPDYPDIYGAIDQNGTYQNVRTGEDGKLTFFKNLSD 796
QY 417 GTYLVVEKKAPLGNLLDNSQ-----KVILGD-----GATDTTNSDNLAVNPT 459
Db 797 GKTRLFENSEPAGYKPVQNKPIVAFQVINGEVRDVTISVPQDIPAGVEFTNDKHYITNEP 856
QY 460 VENNKGTSLPSTGGIGTTFIYIIGAILVIGAGIVLVARR 498
Db 857 IPPKR--EYRPTGGIGMLPFYLGICMM--GGVLLVTRK 891
RESULT 15
B86807
hypothetical protein yoiC [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86807
R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: B86807
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1441 <STO>
A:Cross-references: GB:AE005176; PID:g12724450; PIDN:AAK05556.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: yoiC
Query Match 7.8%; Score 200; DB 2; Length 1441;
Best Local Similarity 23.9%; Pred. No. 0.0045;
Matches 132; Conservative 61; Mismatches 199; Indels 160; Gaps 30;
QY 23 VTPTAFAAETGTT-----VQDTQK---GATYKAYKVPDAEIDNA 59
Db 799 VTPTALAGQNTTDTYSVVQYNTNTIKGISVPVDSGTGAIAKATAIEKYSDETVNA 858
QY 60 NVSDSNKDGASYLIPQGEAEYKAST--DFNSLFTTTNGGRYTVTKDTPASANEIATW- 116
Db 859 PTAP-----AGYVL--NDVNNKAVSVADANGVINFYVDIAIQTLLTKD-STIDEGSTWT 909
QY 117 -AKSISANTTPVST-----VTESNNDGTEVINVSQYGYVYVSVTVNNGAVIMVTSVTPN 169
Db 910 AADNFTGGTSGTGLVLTINDVTVTGTVDTPGVSSTVYTYTDP--TGASI---SSVAN 964
QY 170 ATTHEKN--TDATWDCGGK--VQDKTY-----SVGDTVKYTYTYKNAVNVHGTKE 217
Db 965 ITVNDSSNTNPTDGTGGGNTNPTDGTNIVLPQDGKLTADDTVTGQDPLSPSGDS 1024
QY 218 VYQYVIKDTMPS-ASVVDLNEGSEYV---TITDGSNI-----TTLTQGEKATGKNLL 268
Db 1025 V-----TLPSDGGKVDKPDGSGVNPVPGTVPDPGTIHLPGGTVINPGG----- 1067
QY 269 EENNFTITIPWAATNPTGNTQNGANDPFYKGINTI-----TVTYTGLVKSAGKES 322
Db 1068 -----SVTVP-----GPDGKT--GTDDDTTLNPNSPVPVPGDNGSVTLPGGGTASTPNGN 1114
QY 323 ADLPENTNIATINPNTSNDPQKQVTVRDGQITIKIDGST-----KASLOGAIFVLKN 376
Db 1115 IITLPGGT--VVPDGTIHLPGGDIVNPDGTITLPGDQKGTGDDGKIPNGFIIPGDN 1171
QY 377 AT-----GQFLNFDNNVWEGT--EANAETVTTGA-----DGIITITGLKEGTYYLVE 423
Db 1172 GSVTLPGGGVTTTPGGTINVPGGSVWDPDGTVHLPGGDIVNPDGTITLPG----- 1221
QY 424 KKAPLGYNLLDNSQVILGDGATDTTNSDNLVN-PTVENNKGT-ELPSTGGIGTITFYI 481
```

```
Db 1222 -----QDGKTGTGDDGKVKPNGPSISNPDGSITLPGGGTVTTP----- 1259
QY 482 IGAILVIGAGIV 493
Db 1260 GGTINVPGGSVV 1271
RESULT 16
AH1396
peptidoglycan anchored protein (LPXTG motif) [imported] - Listeria monocytogenes (strain
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH1396
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1396
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1530 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00654.1; PID:g16412064; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2576
Query Match 7.7%; Score 199; DB 2; Length 1530;
Best Local Similarity 21.7%; Pred. No. 0.0055;
Matches 141; Conservative 74; Mismatches 192; Indels 244; Gaps 33;
QY 18 MAVSPVPIAPAAETGTTIVQDTQKATYKAYK-----VFDAEI--DNANVS 62
Db 675 MLITFKKPLIHVE---VTYKTPVGITKPLPKYKAVISDGEVLADYEAEVIDNAN-- 729
QY 63 DSNKDGASYLIPQGE-----AEYKASTDFNSLFTTTNGGRYTVTKDTPAS--- 109
Db 730 -----KYVNSGQVGNIDWEIYANGSGSTVSNATVTDLTGTGQ---KLTSSIKV 778
QY 110 -ANEIATWAKSISANTTPVS-----TVTESNNDGTEV-----INVSYGYVYVSVST 154
Db 779 YKSQTSVTQKMLQESNMPISPGEYDLKTGVDEESLEVPQVFKNEINQSYVIKYQTAIT 838
QY 155 -----VNGGAVIMVTSVTPNATTHEKNTDA--TWDCGG--KTVDQ-- 191
Db 839 LITSDETTHAQIGNSVTFGDNITKGETEKTAKIEVKITTDGDTGTGETKIIILNKVDKAD 898
QY 192 -----KTSYSGDVTKYTYTYTYKNAVNVHGTKEVY-QYVIKD-----TMPASV- 232
Db 899 PSIPLEGATFDLYANDEKVDYDTQTDKNGV-IEFDLLVYGDYTLKEVSAPEGVTLPTASTE 957
QY 233 ---VDLNEGSEYVITD-----GSGNITTLTQGEKATG-----KYNL----- 267
Db 958 NIQVLEODEKVVQVNMNEKMPIKETGEVHLVK--TDKATGATLAGAEFSLYDKSGAELQN 1015
QY 268 ---LEENNFTI---TIPWAATNPTG-----NTQNG-----ANDDF 299
Db 1016 GLTTDENGELTIHNLDLGSYLYKETKAPGKYLSEKTWFEVSQGVDAIEIOAENEKDL 1075
QY 300 YGINTITVTVYTVLKSAGK-----PG-----SADL 325
Db 1076 GEAVLTQVDSKTNKSLGAKFNLLNDSGEVIQTNLVSDENGEIRVQNLEPGGYAFQETEA 1135
QY 326 PENTNIATINPT-----SNDPQKQVTVRDGQITIKIDGSTKASLQ 368
Db 1136 PTNYDIAT---NTWPTTIVAGTSATMTAENNKGT-KPDVDTGEVILVQDSATGETLE 1191
QY 369 GAIFVLKNATGQFLNFDNNVWEGTEANATYTTGADGIITITGLKEGTYYLVEKKAPL 428
Db 1192 GAVFDLMTADGAIV-----ASNLTTDANGEITVTNLPAPGKYSFKETKAPE 1236
```

QY 429 GYNLLNSOKVILGDGATDT---TNSDNL--LVNPTVNNKGTLPSTGGI 474  
Db 1237 GYEL-----ATDWBEPTAPNOPEKITTITAENTKLAIPIDAGSV 1275

RESULT 17  
AII094  
probable peptidoglycan bound protein (LPXTG motif) lmo0160 [imported] - Listeria monocytogenes  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AII094  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, P.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, D.; Jones, L.M.; Karst, U.  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AII094; MUID:21537279; PMID:11679669  
A:Accession: AII094  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-571 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC98375.1; PID:gl6409519; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0160

Query Match 7.5%; Score 192.5; DB 2; Length 571;  
Best Local Similarity 21.4%; Pred. No. 0.0034; Mismatches 191; Indels 191; Gaps 26;  
Matches 122; Conservative 65; Mismatches 191; Indels 191; Gaps 26;

QY 46 YKAYKV-FDAEIDNA-NVSDSNKDGASYLIPQCKEAYKASTDFNSLFTTTTNGGRT--- 100  
Db 64 YDSIKVHYTWEPNSTNVKEG-DTWEFVLP---PELKIVTDLD--FSLKDHGDNVTGCH 115

QY 101 YTKDKDTASANEIATWAKSISANTTTPVSTVTSNNDGTGVINVSQGYVY---VSS 153  
Db 116 VIAKSKTGQV-VITTFDFEKNSS-----NISGYLDFWTNWDKSLVSG 156

QY 154 TVNNGAVIMVTSVTPNATIHKNKTDATGWDGSGKTV--DQKTVSVG-----DTVKYTI 204  
Db 157 NENVPVEFPVNGTGTIDV-----GVGKNQIDPDDESLYKYGWADAENPELIQVWV 207

QY 205 TYKNVAVNYHGTREKVOYVIKDTMPSASVVDLN-----EGSYEV--TITDG-----SGNITT 253  
Db 208 ----RVNY-SKQINQAVVEDFIGPVQVDFNSIKAFHGFEPDDNFTFGAEPSSAITQ 262

QY 254 LTQSEKATGKYNLLBNNTTITIPWAAATNTPTGNTQNGAN-----DDFPYKGINNTIT 307  
Db 263 TTDGFKVDLG--NLTD-----SVKISYTTSTDNGASPNYTKGQLTGDNFIKQIEVAT 315

QY 308 VITYTVGLKSGAKPGSADLPENTNIATNP-----DDPGQKV-----NTS 339  
Db 316 PTSGGGGGEGTGTGSVELTKTDDSSQKNPLEGABFLVNGAGATVGTGLKTWNADGKLATS 375

QY 340 N-----DDPGQKV-----TVRDQGITIKKIDGSTKA 365  
Db 376 NLKYDTYQLIETKAPQGYVLDASPVKFTIDDTTHQSLFVSKENTAIKGSVSLKVDRTKS 435

QY 366 SIQGAIFVLKNATGQFLNFDNTNNVEWTEANATEYTTGADGIITITLKEGYIYLVK 425  
Db 436 LLADAEFELQDKDGNLT-----TNLKTQKGLTVADLLPGEYQVETK 480

QY 426 APLGYNLLNSOKVILGDGATDTTNSDNLNVNPTV-----ENNKGT 467  
Db 481 APTGYILDTTLPFKISTEALNVTVTKENTKPEIKVPVPPKPKPKDKIISDSKQTA 540

QY 468 LPSTGGIGTTFIYGAILVIGAGIVLVA 496  
Db 541 LPKTG-----DSPLVNGWGLLVA 559

RESULT 18  
AB1283  
peptidoglycan linked protein (LPXTG) [imported] - Listeria monocytogenes (strain EGD-e)  
C:Species: Listeria monocytogenes  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AB1283  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, P.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Makok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, D.; Jones, L.M.; Karst, U.  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AB1283  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1711 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC99744.1; PID:gl6411102; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo1666

Query Match 7.4%; Score 191; DB 2; Length 1711;  
Best Local Similarity 21.7%; Pred. No. 0.017; Mismatches 127; Conservative 73; Mismatches 184; Indels 200; Gaps 28;

QY 17 GMAVSPVTPIAFAAETGTTTODTKGATYKAYKVDFDAEIDNANVSDSNKDGASYLIPQ 76  
Db 1226 GNKADPVTI-----TIVVADTEK-----PIITADTTTYAKGTTKTVQAF 1266

QY 77 KEARYKASTDFNSLFTTTTNGGRTYVTKDKTASANEIATWAKSISANTTTPVSTVTSNND 136  
Db 1267 -----TDIHA-----ITNDGSTIISNFPDPTVLAQEGTY-----TVVNAKDESSNE 1307

QY 137 GTEVINVSQGYVYVSTV--NNGAVI-MVTSVTPNATIHKNKTD-----ATWGDGGK 187  
Db 1308 ADPVT-----VTITVVDTKGPIINALNATYERTINKNEADFLADIEATTDGSGTI 1358

QY 188 TVD-----OKTVSGDVTKYTITYKNVNYH-----G 214  
Db 1359 TTFNSKDLDTVGTIVIVLNAEDASGNKATPVKTVIKVEDTIPPIITADQSIYERGITK 1418

QY 215 TEKVYQYVIK-----DTMPSA-----SVVDLNE-GSYEVTI--TDGSGNITLTQSEKATG- 263  
Db 1419 TEQAFVTDIKAATSDNSPISDFSKIDLTKTGNVEVLLRATDSGN-----KALPL 1469

QY 264 KYNLLBNNTTITIPWAAATNTPTGNTQNGANDDPYKGINNTITVITYTVGLKSGAKPGSA 323  
Db 1470 KINVLVQD-----TIAPVKTTSREITA 1492

QY 324 D-----LPENTNIATNPNTSNDPQKVTV-----RDGQITIKKIDGSTKASLQ 369  
Db 1493 ERGTPTWTEOQLAKIGANT---DQSGKITDYNPAIVNTSGDYLVHLYVADAGNQAI 1549

QY 370 AIFV-LKNATGQFLNFDNTNNVEWTEANATEY-----TTGADGIITITG-----LKEGT 418  
Db 1550 EITIHVKDDTAPIIKADKKISYVPVGVTVTFEQLDIIHATTDGSKITTDGPNMLKTPG 1609

QY 419 YVLV-----EKAPLGYNLLNSOKVILGDGATDTTNSDNLV 456  
Db 1610 KYTIHLNVAADVADGNKAKTIDVSLTVEEKVTPPKPTDDGGNNTNGSNGTSNTD-ITV 1668

QY 457 NPTVNNKGT--ELPSTGGIGTTFIYGAILVIGAGIVLVARR 498  
Db 1669 NPTKQNTATNESIPALGDTKSTIPVIGMFL-LATSLVLRK 1711

RESULT 19  
S41525  
major ring-forming surface protein precursor - Helicobacter mustelae  
C:Species: Helicobacter mustelae

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 28-May-1999	
C;Accession: S41525; S77974	
R;O'Toole, P.W.; Austin, J.W.; Trust, T.J.	
Mol. Microbiol. 11, 349-361, 1994	
A;Title: Identification and molecular characterization of a major ring-forming surface p	
A;Reference number: S41525; MUID:94224153; PMID:8170397	
A;Accession: S41525	
A;Molecule type: DNA	
A;Residues: 1-1519 <OTO>	
A;Cross-references: GB:L15629; NID:g431303; PIDN:AAC36865.1; PID:g431304	
A;Accession: S77974	
A;Molecule type: protein	
A;Residues: "XX",50-59 <AUS>	
C;Genetics:	
A;Gene: hsr	
F;1-47/Domain: signal sequence #status predicted <SIG>	
F;48-1519/Product: major ring-forming surface protein #status experimental <MAT>	
Query Match 7.3%; Score 189; DB 2; Length 1519;	
Best Local Similarity 22.9%; Pred. No. 0.019;	
Matches 125; Conservative 70; Mismatches 206; Indels 146; Gaps 27;	
Qy	3 KQMISLLVSLAFGMVSPVPIAFAAETGTTTVDQTKGATYKAYKVPDAIDNANVS 62
Db	25 RKFPQPIATTLAFSLSPFN---AADAG-----NAGQA 56
Qy	63 DSNKDGASYLIPQKEAEYKASTDFNSLFTTTNGGRTYVTKDTASANEIATWAKSISA 122
Db	57 PVNAEGITVTNQANKATATVSGNNGNATF-TFTNGANT--TVNGTA---DPAVTAPEV 110
Qy	123 NTPVSTVSTESNNDGTEVINVSQ-----YGYVVSS-----TVNNGAVIMV 163
Db	111 NI--ANIVNFTVDCKPANQANQNLGAEGPVNLFNPFGLIASSGTAKTFTLNLGGAGNA 168
Qy	164 TSVT-----PNATHEK-----NTDATWGD--GGKTVDDQKTYSGVDT 199
Db	169 NALTGNLNLGAGNATLNTNGSIASGGPVINVNKDATFNATFSGGATMTGNIVT-GNT 227
Qy	200 VKYITVKNVNYHGTKEVQ--VVIKD-----TMPASVVDLNEGSY----EVTITDG 247
Db	228 KETSGTGNNITFDGPKQIPHNGSLIKDGTAVTQADPAVLTGNISTYGINNVTFEK 287
Qy	248 -----SGNITTLTQSEKATGYNLLENNNFTIPWAAATNTPGNTQNGANDDF-F 299
Db	288 TMKGDIAGNATQSLGNVVTFK---EQGVHYTGNVIASTGT-GVNNITLNFGNATVDA 342
Qy	300 YKGINITVTVTGVLKSGAKPGSADLPENTINIATINENTSNDDPGQKVTYVDQGIITIKI 359
Db	343 TNGGNLTLIQNSGITFNNTN-GVNNSTPLTH-AITTPAAAGGDPANQATVPQGNI----- 395
Qy	360 DGSTKASI-QGAI FVLKNATGFLNF-----NDTNVWEGTEANATEVTTGADGIITITGL 414
Db	396 -----KSAIQGV-----NTLNFYFAKLEGTTPANKANPAPAANITATNNGANNIVFTDG- 444
Qy	415 KEGTYVLEKAPLGYNLLDSQKVLGDGA-TDTTNSDNLNVP-TVENNKGTELPSTG 472
Db	445 -----GLVNAALTSLDQGINLTVMNTNIVTNPIILLTGNNVNTTFCWA 488
Qy	473 GIGTITP 479
Db	489 GSNLTLLF 495
RESULT 20	
T33369	
hypothetical protein H02F09.3 - Caenorhabditis elegans	
C;Species: Caenorhabditis elegans	
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000	
C;Accession: T33369	
R;Geisel, C.; Harmon, G.	
submitted to the EMBL Data Library, July 1998	
A;Description: The sequence of C. elegans cosmid H02F09.	
A;Reference number: Z21330	
A;Accession: T33369	
A;Status: preliminary; translated from GB/EMBL/DBJ	
A;Molecule type: DNA	
A;Residues: 1-1275 <GEI>	
A;Cross-references: EMBL:AF077538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3	
A;Experimental source: strain Bristol N2; clone H02F09	
C;Genetics:	
A;Gene: CESP:H02F09.3	
A;Map position: X	
A;Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1	
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase 1	
Query Match 7.2%; Score 186.5; DB 2; Length 1275;	
Best Local Similarity 21.4%; Pred. No. 0.02;	
Matches 114; Conservative 86; Mismatches 178; Indels 155; Gaps 24;	
Qy	17 GNAVSPVT-----PIAFAAETGTTTVDQTKGATYKAYKVPDAIDNANVS-DSNKGAS 70
Db	576 GSSISTVSTVSSQFSTYIPVSSASSIYTLSGSTASTASPGTTBSSGSSSTGSPSTISGSS 635
Qy	71 YLIPQKEAEYKASTDFNSLFTTTNGG-RTYVTKDTASANEIATWAKSISANTTPVST 129
Db	636 ASTVTGSIIVT-EASTISGSTESTIPGSTESTVSEASTVSGSSVST-----VSG 683
Qy	130 VTESNNDGTEVINVSQYGYVVSVTVNNGAVIMVTSVTPNATIHKNTDATWGDGGKTV 189
Db	684 STESTSAGASTVSGS-----TGSTVSDSTISDSTGSTNA---PGTESTVTGSSSVTV 734
Qy	190 DKTYTVSGDTVKYITVKNVNYHGTKEVQYVVKDTMPSASVVDLNEGSYEVIITDGS 249
Db	735 SGSTGSTGPS-----TMSASTGNTTGTSTESTITDGS- 767
Qy	250 NITTLTQSEKATGYNLLENNNFTI---TIPWAAATN---PTGNTQNGANDDFYKGI 303
Db	768 ----TVSGSTGSGTNNPGSDSTTGTISVSGSSLSTISGSGSTVSGSSD----- 816
Qy	304 NITVTYTVGLKSGAKPGSAD-----LPENTINIATINENTSNDDPG 344
Db	817 --MTVS-TG---STSPGSGSTVSGASTMPSGSTSGSVSTSGSVSTSGSVSTSGSTG 870
Qy	345 Q-----KVTVRDGOITTKIDGSKASLQGAIF-----VLKNATGQFLNFNDT 387
Db	871 QSTVSESSVSTVSSSESTISQSTGTTG-ESIVFGSTGSTATGSTMSTGSGS---TDT 925
Qy	388 -----NNVWEGTEANATEYTTGADGIITITGLKEGTYVLEKAPLGVN 431
Db	926 PGSTESTITGSTVGTGESTVSGSTGSGTITGSGTISESTMTTVGVSTGCTITGES-----T 979
Qy	432 LLDNSQKVLGD---GATDTTNSDNLNVPVNNKGTFLPST--GGIGTTI 478
Db	980 VSGSTRSTVGTGESTVSGSTESTVSGSTESTPTIV-----PSTVSGSTGTV 1024
RESULT 21	
AD0123	
probable autotransporter protein yaph [imported] - Yersinia pestis (strain CO92)	
C;Species: Yersinia pestis	
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001	
C;Accession: AD0123	
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;	
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;	
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, H.	
Nature 413, 523-527, 2001	
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.	
A;Reference number: AB0001; MUID:21470413; PMID:11586360	
A;Accession: AD0123	
A;Status: preliminary	
A;Molecule type: DNA	
A;Residues: 1-3705 <KUR>	
A;Cross-references: GB:AL590842; PIDN:CAC89847.1; PID:gl15979073; GSPDB:GN00175	
C;Genetics:	
A;Gene: yaph	



Query Match		7.2%; Score 186; DB 2; Length 3705;
Best Local Similarity		23.5%; Pred. No. 0.088;
Matches 112; Conservative		70; Mismatches 213; Indels 82; Gaps 21;
QY	33	GTITVQDTQKGATYK---AYKVFDAEID--NANVSDSNKOGASYLIPQKAEAYKASDT 86
DB	473	GIINTSATANGITPAGTEGHLTDLTINLGTGIALSNVAGUNLTL-----SNT 523
QY	87	FNSLFTTTTNG--GRITYVTKDTASANEIATWAKSISANTTPVSTVTESSNNDGTEVINVS 144
DB	524	LNTLNGTALNSLTGLTLDVSLNGRNTINIEGAGIGIAATNTELTNTEFAEALD----INVN 579
QY	145	QVGYVYVSVTVANGAVIMVTSVTPNATHEKNTDAT---WGDGGKTVQKYSVGDTVK 201
DB	580	GAG---IGIQTGGGVNLSAS---NLIINVANTLGTALQITDG-----IDNTT-TIGNEIQ 628
QY	202	YTITYKNAVNYHGTEKV---YQVVIKDTMPSASVVD---LNEGSEVTTITDGSNNITLT 255
DB	629	LNAENATAINFLOGSSKTLNNGTIKGSVIFAGVADHIIINNGTLDGLTITGAGNDTLVL 688
QY	256	QGSEKATGKYNLLBENNFTIIPWAATNTPTGNTONGANDDPFYKGINTIITVYTGVLK 315
DB	689	DSSQSNDVINLGDGNSVTIQNGATVSSIIIGN---GNDITFTNGM-SVGSTYLGSLD 743
QY	316	SGAKPGSADLPENTNATINPTNSDDPGQKTVR-----DGOITIKKID--GSTKA 365
DB	744	AG-----TGLNTLNFNASTDELAASLOQFTNINLVDSHITLVSDDNIGSGMV 792
QY	366	SLQGAIFVLKNATQGLNFNDTNVWEGTEANATEYTTGADGIIITITGLKEGYVYLVEKK 425
DB	793	NIDSSSELLFGST-----FDGILHATLGAGTSGAIVNNSANVSLEQASMPAGTQWNOGG 847
QY	426	APLGYNLLDNSOKVILGD---GATDTTNSDNLVNPTNVENNKGTLPSTGGIGTTF 479
DB	848	A-----LTASNSQLGSAKIGLDGTLNLDNLIAFNHVLGTNGT-LNVAKNLATTAF 897
RESULT 22		
S57721		
CpB protein - Clostridium acetobutylicum		
C:Species: Clostridium acetobutylicum		
C:Date: 19-Oct-1995 #sequence_revision 17-Nov-1995 #text_change 15-Oct-1999		
C:Accession: S57721		
R:Sanchez-Beato, A.; Garcia, J.		
submitted to the EMBL Data Library, July 1995		
A:Description: Molecular characterization of a family of choline-binding proteins of Clo		
A:Reference number: S57714		
A:Accession: S57721		
A>Status: preliminary		
A:Molecule type: DNA		
A:Residues: 1-583 <SAN>		
A:Cross-references: EMBL:Z50009; NID:g895758; PIDN:CAA90304.1; PID:g895759		
C:Superfamily: cpl repeat homology		
F:504-523/Domain: cpl repeat homology <CP2>		
F:544-563/Domain: cpl repeat homology <CP3>		
Query Match		7.2%; Score 185.5; DB 2; Length 583;
Best Local Similarity		21.6%; Pred. No. 0.0082;
Matches 132; Conservative		83; Mismatches 215; Indels 181; Gaps 28;
QY	1	MKKKMIQ---SLIVASLAFGMVSPVTPPIAFAETGTTVQDTQKGATYK--AYK-----VF 52
DB	1	MIKRMTKATSLVAA-----AAIISIVP-----AHAADYTKIDSQEGTIYDAVAYKDGKFFV 52
QY	53	DAEIDNANVSDSNKOGASYLIPQK-----EAEYKASTDF-----NSLFTTTTNG 97
DB	53	DGEV-----NDKDEAAAYLADGKYNLSIDSGADTDIYGSKYLEVODGYFDVLDNG 105
QY	98	GRITYVTKDTASANEIATWAKSISANTTP--VSTVTESSND--GTEV-----INVSOYGY 149
DB	106	SVTDESIVENAEDDAASALRKNLKNDKRYASSERETIQLAGAIAGNKSPAPKYFE 165
QY	150	YVSSVTNNGAV-----IMVT-----SVTPNATHEKNTDATW-GD 183

DB	166	YTKDNATNGSVAKLNVYTDAGQNYIDADYNLGSIKVTTTASSTTSKTVTIANTDSDYDAD 225
QY	184	GGGKTVQDKTVSGVDTVKYTIYKNAVNYHGTEKVYQVYVVKDTMPSASVVDLN-----E 237
DB	226	GTSDATGKKSSASVDTSGKVIQDANYIYRTAK-----VTVNAANGATITKINGMDVTVD 281
QY	238	GSYEVTITDGSNITITLQSGSEKATGKYNLLBENNFTIIPWAATN-----TP 286
DB	282	GGHTFTVSN-----TLGSSASTVGFNVIOK-----ISKAQASGNVDGAKYAKSVTTY 329
QY	287	TGNTONGANDDPFY-----KGINTIIVYITGVLSKSG-----AKPGSADL 325
DB	330	IVSDEDEGRNEAFAYDNYTVANGKIVGYTANGTNVNTATGTLSSKNGYVYVDLGDSEADV 389
QY	326	PENTNIATINPT-----SNDDPGQKVTVRDG---QITIKKIDGSTKASL 367
DB	390	KENGTKSAVDTDVDCWRLDAGYIYKWDNDEDTWKYKVDGSPDOMSVYNQDNIVAWSK 449
QY	368	QGAIFVLKNATQGLNFNDTNVWEG---TEANATEYTTG-----A 405
DB	450	EDVYSVIGGKGTSTPTDPTPVNKGWKTDAGMTFYNTDPTQVKQWVNDGQVWYIYKA 509
QY	406	DGIITITGLKEGT--YVLVEKKAPLYNLLDNSQKVL-GDCATDT-----TN 450
DB	510	DGTMATGIDGSTWYLYLQSSGAMKTGWLNDNGTWYLYLQSSGAMSTGWLNDNGTWYFLNS 569
QY	451	SDNLLVNPTVE 461
DB	570	SGMLANTVTD 580
RESULT 23		
A48995		
paracrystalline surface layer protein RsaA - Caulobacter crescentus		
C:Species: Caulobacter crescentus		
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000		
C:Accession: A48995		
R:Gilchrist, A.; Fisher, J.A.; Smit, J.		
Can. J. Microbiol. 38, 193-202, 1992		
A:Title: Nucleotide sequence analysis of the gene encoding the Caulobacter crescentus pa		
A:Reference number: A48995; MUID:93007489; PMID:1393820		
A:Accession: A48995		
A>Status: preliminary		
A:Molecule type: nucleic acid		
A:Residues: 1-1026 <GIL>		
A:Cross-references: GB:AF062345; GB:M22663; GB:M84760; NID:g6064104; PIDN:AAC38665.2; PI		
A:Experimental source: CB15A, ATCC 19089		
A>Note: sequence inconsistent with the nucleotide translation		
A>Note: sequence extracted from NCBI backbone (NCBI:116173, NCBI:116174)		
Query Match		7.2%; Score 185.5; DB 2; Length 1026;
Best Local Similarity		22.1%; Pred. No. 0.017;
Matches 131; Conservative		66; Mismatches 240; Indels 157; Gaps 26;
QY	6	IQSLVLASLAFGMVSPVTPPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSN 65
DB	299	VQAAAVTALPTGVTISGIE--TMNVTSGAATLNTSSGVT-----GLTALNTN 344
QY	66	KDGASYLIPQKAEAYKASTDFNSLFTTTTNG-----RTVYTK-KDTASANEIATWAK 118
DB	345	TSGAAQTVTAGAGQNLTLATTAQAANNVAVDGGANNVAVASTGVTGTTTGGANSASGTV 404
QY	119	SIS-ANTTIPVST-----VTESNDCGTEVINVSQYV 148
DB	405	SVSVANSTTTTCAVATGCTGTAATVTAQTAGNAVNTLTQADVTVTGNSSTTAVTQTAA 464
QY	149	YVSSVTN-----NGAVIMVTSVTPNATHEKNTDATWGDGGKTVQKYSVGDTVKTYIT 205
DB	465	ATAGATVAGRVNGAVTITDSAAASATTAGKIATVTLGSFGAATIDSSALT----- 514
QY	206	YQNAVNYHGTEK---VYQVVIKDTMPSASVVDLNGSEVTE--ITDGG-----NITLTQ 256

Db 515 ---TVNLSTGCTSLIGRGALAT-PTANTLTNLVNGLTITGTAITDSEAAADGGFTTINI 570  
Qy 257 GSEKATGYNLLBENNFTTITIPWAATNTPTGNTQNGANDDPFYKGINITVT----- 309  
Db 571 AGSTASSTIASLVAADATLTNISGDARVITTSHTA-----AALTGITVTSNGATL 621  
Qy 310 ----YTCVLKSGAKPGSADLPENTNIA-----TINPNTSNDPQKQVTVRDG-QITI 356  
Db 622 GAELATGLVFTGGAGADISLLGATTKAIVMGAGDDTVTVSSATLGAGGSGVNGDGTDLV 681  
Qy 357 KKIDGSTKASLOGAIFVLKNATGQF-----LNFDNTNNVWEGTEANATEYTT 403  
Db 682 ANVNGSS-----FSADPAFGGFETLRVAGAAQOSHANGFTALQCATAGATTFTN 733  
Qy 404 GADGI-ITITGLKEGTYYLVEKAPLG-----YNL-LDNSQKVLGD-----GAT 446  
Db 734 VAVNVGLTVLAAPTGTITVTTLANA-TGTSDFVNLTLSSSAALAAGTVALAGVETVNIAT 792  
Qy 447 DTTNS---DNLLVNPT-----VENNKGTLPSTGGIGTTFIYIIGAILVIGAG 491  
Db 793 DTNTTAHVDTLTLOATSAKSIVVTGNAGLNTNTGNTAVTSF---DASAVTGTG 843  
RESULT 24  
C87374  
S-layer protein ReaA [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: C87374  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: C87374  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1073 <STO>  
A:Cross-references: GB:AE005673; NID:g13422297; PIDN:AAK22991.1; GSPDB:GN00148  
A:Genetics:  
A:Gene: CC1007  
Query Match 7.2%; Score 185.5; DB 2; Length 1073;  
Best Local Similarity 22.1%; Pred. No. 0.018;  
Matches 131; Conservative 66; Mismatches 240; Indels 157; Gaps 26;  
Qy 6 IQSLLVASLAFGMVSPVTPPIAFAETGTTIVQDTQKATYKAYKVPDAEIDNANVSDSN 65  
Db 346 VQAAAVTALPTGVTISGIE--TMNVTSAAITLNTSSGVT-----GLTALNTN 391  
Qy 66 KDGASVLPQKKEAEYKASTDFNSLFTTTTNGG-----RTYVTK-KDTASANEIATWAK 118  
Db 392 TSGAAQVTTAGAQNLTAATAQAANNVAVDGGANVTVASTGVTSGITTVGANSAAAGTV 451  
Qy 119 SIS-ANTTPVST-----VTESNNDGTEVINVSQY 148  
Db 452 SVSVANSSTTTTGALAVTGGTAVTVAQTAGNAVNTTLQADVTVTGNSSTTAAVTQTAA 511  
Qy 149 YVVSSTVN---NGAVIMVTSVTPNATIEKNTDATWGGGKTVTDQKTSYSGVDIVKYIT 205  
Db 512 ATAGATVAGRVNGAVTITDSAAASATAGKIATVTLGSGAATIDSSALT----- 561  
Qy 206 YKNAVNYHGTGK---YVQVVKDTMPSASVVDLNEGSVEVT---ITDGGG---NITTLTQ 256  
Db 562 ----TVNLSTGTSLSIGRGALAT-PTANTLTNLVNGLTITGTAITDSEAAADGGFTTINI 617  
Qy 257 GSEKATGYNLLBENNFTTITIPWAATNTPTGNTQNGANDDPFYKGINITVT----- 309  
Db 618 AGSTASSTIASLVAADATLTNISGDARVITTSHTA-----AALTGITVTSNGATL 668  
Qy 310 ----YTCVLKSGAKPGSADLPENTNIA-----TINPNTSNDPQKQVTVRDG-QITI 356

Db 669 GAELATGLVFTGGAGADISLLGATTKAIVMGAGDDTVTVSSATLGAGGSGVNGDGTDLV 728  
Qy 357 KKIDGSTKASLOGAIFVLKNATGQF-----LNFDNTNNVWEGTEANATEYTT 403  
Db 729 ANVNGSS-----FSADPAFGGFETLRVAGAAQOSHANGFTALQCATAGATTFTN 780  
Qy 404 GADGI-ITITGLKEGTYYLVEKAPLG-----YNL-LDNSQKVLGD-----GAT 446  
Db 781 VAVNVGLTVLAAPTGTITVTTLANA-TGTSDFVNLTLSSSAALAAGTVALAGVETVNIAT 839  
Qy 447 DTTNS---DNLLVNPT-----VENNKGTLPSTGGIGTTFIYIIGAILVIGAG 491  
Db 840 DTNTTAHVDTLTLOATSAKSIVVTGNAGLNTNTGNTAVTSF---DASAVTGTG 890  
RESULT 25  
A83412  
hypotheetical protein PA1874 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83412  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Miziochuchi, S.D.; Warrenner, P.; Hickey, M.J.; Br.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83412  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2468 <STO>  
A:Cross-references: GB:AE004613; GB:AE004091; NID:g9947856; PIDN:AAG05263.1; GSPDB:GN0013  
A:Experimental source: strain PA01  
A:Genetics:  
A:Gene: PA1874  
Query Match 7.2%; Score 184.5; DB 2; Length 2468;  
Best Local Similarity 21.0%; Pred. No. 0.062;  
Matches 116; Conservative 69; Mismatches 209; Indels 159; Gaps 24;  
Qy 21 SPVTPIAFAAEATGTI---TVQDTQKATYKAYKVFDA-----EIDN-ANVSDSNKDG 68  
Db 1044 TFSFPLA---DGTVVNATATDPAGNTGGGSTTVDAIAPATPTVNLNSGSSLSGTAEPPG 1099  
Qy 69 ASYLIPOGKEAEYKASTDFNSLFTTTTNGR---TYVTKDITASANEIATWAKSISANTTP 126  
Db 1100 STVILTGD-----NGNPIAEVTADGSGNWTYTPTPIANGTVVNVVVAQDASGNSP 1150  
Qy 127 VSTVT-ESNNDGTEVINVSQYGYVVSSTVNVNGAVIMVTSVTPNATIEKNTDATWGDG 185  
Db 1151 PATVTVDSGAPPAPVINPS-----NGVVISGTA-----EAGATVTLTDAG 1190  
Qy 186 GKTVDQKT-----YSVGDTV-KYTIITYKNAVNYHGTGKVVYQYVVKDTM-PSASVVDL 235  
Db 1191 GNPIGQVTTADGSGNWSFTPTGPLANGTVIVATADTPTGNTGPOAATTVDAVAPPAVIDP 1250  
Qy 236 NEGS-----YEVITDGSNIT-----TLTQSEKATGYNLLENNFT 275  
Db 1251 SNGTTSIGTAEAGAKVILTNGNPNIGETTADGSGNWSFTPGPLANG----- 1298  
Qy 276 ITIPWAATNTPTGNT---QNGANDDFYKGINITTVTVTVGLKSGAKPGS----- 322  
Db 1299 -TVNVAQDPAGNTGPGQSTTVDAVAPNTPVVPSNGNLLNGTAEPGSVTLTDCGNP 1357  
Qy 323 -----ADLPENTNI-ATINPNTSNDPQKQVTVRDGQITIKIDGSTK 364  
Db 1358 IGQTTADGSGNWSFTPGSQLPNCVTNVNVTASDAAGNTSLPATTTVDSLSLPSIPQVDPNSG 1417  
Qy 365 ASLQGAIFVLKNATGQFLNFDNTNNVWEGTEANAT---EVTTCADGIIITIT---GLKEGT 418  
Db 1418 SVISGTA-----DAGNTIIITDGNPNIGQVTTADGSGNWSFTPGIPLPDGT 1463  
Qy 419 YVLVEKAPLGYNLNDSQKVLGDGATDTTNSDNLNVNPTVNNKGTTELPTSGGIGTTI 478

Db 1464 VNVVARSF---SNVDSAPAVITVDGVAPAA-----PVIDPSNGTISGTAERAGATV 1512  
QY 479 FYIIGAILVIGAG 491  
Db 1513 -----ILTDGGG 1519

RESULT 26  
T28679  
fibrinogen-binding protein homolog - Staphylococcus aureus  
C:Species: Staphylococcus aureus  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C:Accession: T28679  
R:Josefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.  
Microbiology 144, 3387-3395, 1998  
A:Title: Three new members of the serine-aspartate repeat protein multigene family of St  
A:Reference number: Z20510; MUID:99098700; PMID:9884231  
A:Accession: T28679  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1315 <JOS>  
A:Cross-references: EMBL:AJ005646; NID:e1318791; PID:e1318792; PIDN:CAA06651.1  
C:Genetics:  
A:Gene: sdR

Query Match 7.0%; Score 181.5; DB 2; Length 1315;  
Best Local Similarity 22.0%; Pred. No. 0.039;  
Matches 116; Conservative 72; Mismatches 191; Indels 149; Gaps 29;

QY 34 TITVQDTQKATYKAYKV--FPAEDINAVNSDKGAS---YLIPQKEAE--YKAST- 85  
Db 427 TIYVNPSENSLNAKLKVQAYHSSYPN-NIGQINKDVTDIKYQVPKGYTLNKGYDVNTK 485  
QY 86 -----DENSLETTTNGGRTVV-----TKDKTASANEIATWA--- 117  
Db 486 ELTDTVNQYLKITYGDNNSAVIDFGNDSAYVMWNTKFQYTNSESPLVQMATLSSTG 545  
QY 118 -KSISANTTPVSVTBSNNDGTEVINVSQYGYVSVSTVNNGAVIM---VTSVTPNATI 172  
Db 546 NKSVSFGNALGFTNNGSGGAGQEVYKI---GNVWEDTNKNGVQELGEGKGVNVT--VTV 600  
QY 173 HEKNTDATWGDGGKTVQDKTYSV-----GDTVKYITIYKN----- 208  
Db 601 FDNNTNWKVGE--AVTKEDGSLIPLNPNLPGD---YRVEFSNLPKGYEVTTPSKQGNNEILD 655  
QY 209 -----AVNYHGTKEK---VYQ-----YVQKDTMPSASVDLNE---GSYEVTITD 246  
Db 656 SNGLSVITVNGKDNLSADLGHYKPKYNLGDYVWEDTNKN-GIQDQDEKIGISGVTVTLKD 714  
QY 247 GSGNITFTLQGSSEKATGKYNLLE-ENNFTIITIPWAAATNPTPTGNTQNGANDDPFYKGINT 305  
Db 715 ENGNV--LKTVTTDADGKYKFTDLONGYKVEFTTPEGYTPTVT--SGSDIEKDSNGLTT 771  
QY 306 IIVTVTGVLKSKAGPSADLPENTNIATNPNTSNDNDPQKV---TVRDGQITIKIDGS 362  
Db 772 -----TGVI--NGA-----DNMTLDSGFYKTPKYNLGNYYWEDTNKDGK-----QDS 811  
QY 363 TKASLOGAIFVLKNATGQFLNFDNTNVEWTEANATEYTTGADGIITITGLKEGYTLV 422  
Db 812 TEKIGISGVTVTLKNGENGEVLQTTKTDK-----DGKYQDTGLENGT-YKV 854  
QY 423 EKKAPLGYNLLDNSQKVLGDGATDPTNSDNLNVPVNNKGTLPSS 470  
Db 855 EFETPSGYT-----PTQVSGTGDEGIDNSNGTSTTGVIKDKNDRTDS 896

RESULT 27  
S76109  
hypotheical protein - Synechocystis sp. (strain PCC 6803)  
C:Species: Synechocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999

C:Accession: S76109  
R:kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S76109  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3029 <KAN>  
A:Cross-references: EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BAA10087.1; PID:d1010732  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 7.0%; Score 181; DB 2; Length 3029;  
Best Local Similarity 20.5%; Pred. No. 0.13;  
Matches 117; Conservative 63; Mismatches 199; Indels 192; Gaps 26;

QY 20 VSP-VTPIAFAPAEETGTTITVQDTQKATYKAYKVFAEDINAVNSD-----SNKGASYL 73  
Db 580 IAPIVTSILRNPTSEITNADS---LTFQV--TFSENQVNVNDVDFILSGTGVSGATISN 634  
QY 74 PQGKEAEYKASDFRNSLFTTTNG--GRYTVTKDTASANEIATWAKSISANTTP----- 126  
Db 635 VSGS-----GSVTIVTVTGIANNGTVNLDFASGQNIIRDANNAALSNTTPTTDEQ 684  
QY 127 -----VSTVBSNNDGTEVINVSQYGYVSVSTVNNGAVIMVTSV-----TP 168  
Db 685 YTLDNTPAASITLDANITADDDIINAE-----SGQAIPTGTVGGEFNVGDIV 733  
QY 169 NATIHEKNTDATWGDGG-----GKTVQDKTYSV 196  
Db 734 TLTVDNKTFTGAVGAGGLFSINVPGSDLIVDADLTAAASIAATTAAGNLGSATDNQTYTV 793  
QY 197 GDTVKYITIYKNAVNYHGTKEVQYVVIKDWPSASVDLNEGSYEVTTD-----GSGN 250  
Db 794 DTTAPIITVNDVT-----ADNIIAASGQAIPTGTVGGEFNVGD 836  
QY 251 ITTLTQGSSEKATGKYNLLENNFTIITIPWAAATNPTGNTQNGANDDPFYKGINTITVTV 310  
Db 837 TVTLTVNGKPTFG---TVDANGDFSIDV-----LGDLVNGSD-----LTIAA 876  
QY 311 T-GVLKSKAGPSADLPENTNIATNPNTN-TSN-----DD-----PGQKTVTRD 351  
Db 877 SVATTTDAAGNPGSADNCTYTVDTTAPTVTINAIIVADIIINAVEAGSPVAVSGTTTGVED 936  
QY 352 GQITIKKIDGST-KASLOGAIFVLKNATGQFLNFDNTNVE-----MGTEANATEYTTG 404  
Db 937 GQVTVTIDGNVTYATVTGNATFNIPVADIANFEATEEVATVSDLAGNPATPATRNIT 996  
QY 405 ADGI---ITITGLKEGYTLVVEKKAPLGYNLLDNSQKVLGDGATD-TTNSDNLNVPV 460  
Db 997 VDTVAPAVTIDSISDDT-----GAQANDFITNDTLLVFNCTA 1033  
QY 461 ENNKGTETLPSTGGIGTFTTPIIGAILVIGAG 491  
Db 1034 EAD-STVVVSLDGE-----IGTVANGAG 1057

RESULT 28  
T28680  
fibrinogen-binding protein homolog - Staphylococcus aureus  
C:Species: Staphylococcus aureus  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C:Accession: T28680  
R:Josefsson, E.; McCrea, K.; Ni Eidhin, D.; O'Connell, D.; Cox, J.; Hook, M.; Foster, T.  
Microbiology 144, 3387-3395, 1998  
A:Title: Three new members of the serine-aspartate repeat protein multigene family of St  
A:Reference number: Z20510; MUID:99098700; PMID:9884231  
A:Accession: T28680  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1166 <JOS>

A;Cross-references: EMBL:AJ005647; NID:ei1318793; PID:ei1318794; PIDN:CAA06652.1  
C;Genetics:  
A;Gene: sdrE

Query Match 7.0%; Score 180; DB 2; Length 1166;  
Best Local Similarity 23.3%; Pred. No. 0.04;  
Matches 115; Conservative 70; Mismatches 185; Indels 124; Gaps 26;

QY 35 ITVQDTOKGATYKAYKVFDAEIDN-ANVSDSNKDGASYLIPQGEAE-YKASTD-----F 87  
DB 452 IYVNPFLKASATNTKVDITAGSDVDYGNKILGN---GSTIIDQNTIEIKVYKSDQQLPOS 508  
QY 88 NSLFTTTTNGGRTVTKKDTASANEIAT-----WAKSISANTTPVSTVSTESNNDGT 138  
DB 509 NRIYDFSOYEDVTSQPNKKSFSNNVATLDFGINSAYIIKVKVSKYPTSDGELDLTAQGT 568  
QY 139 EVINVSQGYGYYVSSVYNNGAVIMVTSVTPNATIEHKNTDATWGDGGGKTVDQKTSYSGD 198  
DB 569 SMRTTDKYGYYNVAGYSN---FIVTS-----NDTGGDGTGVPK-BEKLKIGD 612  
QY 199 TVKTYITYKNAVYHGTEKYQYVYKDTMPSASVVDLNEGSYEVTTIDGSGNITTLTQGS 258  
DB 613 YWVEDVDKGVQGTDSKEK-----PMANVL-----VTLFYPDG--TTKSVRT 652  
QY 259 EKATGKYNL--LEENNFTITIPWAATNTPTG---NTQNGANDDFYKGINITITVTVGV 313  
DB 653 D-ANGHVEFGGLKDEGYTVKF-----ETPTGYLPTKVNGTTDGERKDSNGSVTVKINGK 706  
QY 314 LKSGAKPGSADLP-----ENVTIATINPNTSNDPQGVTVRDPGQITIK-----KID 360  
DB 707 DMSLDTGFYKEPKYNLGDYVWEDTNKDGIGQ---DANEFG---IKDKVTLKDSGTGKI 759  
QY 361 GSTKASLQGAIFVLKATQFLNFNDTN---VEWGTER-----NATEYTTGADGII 409  
DB 760 GTTTT-----DASGKY-KPTDLONGYVTEPEPAGYPTPVKNITADDKDSNGLT 808  
QY 410 TITLKEGTYLYVEK-----KAPLGYNLLDNSQKVLGDGATDTTNSDNLNVPTVENN 463  
DB 809 TTGVIKDADNWLDRGFYKPKYSLG DYVWYDSNK-----DGKQDSIEKIGKIDVTVLQNE 864  
QY 464 KGTELPSTGGIGTT 477  
DB 865 KGEV-----IGTT 872

RESULT 29  
T28681  
rib protein - Streptococcus agalactiae  
C;Species: Streptococcus agalactiae  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000  
C;Accession: T28681  
R;Wastfelt, M.; Stalhammar-Carllemalm, M.; Delisse, A.M.; Cabezon, T.; Lindahl, G.  
J. Biol. Chem. 271, 18892-18897, 1996  
A;Title: Identification of a family of streptococcal surface proteins with extremely rep  
A;Reference number: 220511; MUID:96324974; PMID:8702550  
A;Accession: T28681  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1231 <WAS>  
A;Cross-references: EMBL:U50333; NID:gl620647; PID:gl620648; PIDN:AAC44468.1  
C;Genetics:  
A;Note: rib

Query Match 7.0%; Score 180; DB 2; Length 1231;  
Best Local Similarity 21.9%; Pred. No. 0.043;  
Matches 127; Conservative 70; Mismatches 200; Indels 183; Gaps 31;

QY 13 SLAFGMVSPVPTAFAAETGTTIVQDTOKGATYKAYKVFDAEID-----NANV 61  
DB 740 TVAFETPVDATPGDKPAKV-VVYTPDGSKDTVDVTVKVVDPRTDADKNDPAGKQDVNV 798  
QY 62 SDSNK--DGASYL--IFQGEAEYKASTDFNSLFTTTTNGR-----TYVT-KKDT--- 107

DB 799 GETPKAEDSIGNLPDLPGKTTVAFETPVD-----TATPGDKPAKVVVVTPDGSKDTVDV 852  
QY 108 -----ASANELIATWAKSISANTTPVSTVTSN-----NDGTEV-----INTVSOYCY 148  
DB 853 TVKVVDPRTDADKNDPAGKQDVNVGNETPKAEDSIGNLPDLPGKTTVAFETPVDATPG- 911  
QY 149 YVSSVNVNGAVIMVT-----SVTPNATIEHKNTDATWGDGGGKTVDQKTSYSGDVTVK 201  
DB 912 -----DKPAKVVVVTPDGSKDTVDVTVKVVDPRTDADKNDPAGK--DQOV-NVGETPK 961  
QY 202 YITYTKNAVY-HGTEKYQYVYKDTMPS---ASVVDLNEGSYEVTTIDGSGNITTLTQGS 257  
DB 962 AEDSIGNLPDLPGKTTVAFETPVDATPGDKPAKV-----VTVPDGSKDTVDVTV- 1011  
QY 258 SEKATGKYNLLENNFTITIPWAATNTPTGNTONGANDDFYKGINITITVTVGLKSG 317  
DB 1012 -----VKVVDPRTD-----ADKNDPAGKQDQ-----VNVG 1036  
QY 318 AKPGS-----ADLPENTNIATINPNTSNDPQGVK-----VTVRDG-----QITIKKID 360  
DB 1037 ETPKAEDSIGNLPDLPGKTTVAFETP-VDTATPGDKPAKVVVVTPDGSKDTVDVTVKVV 1095  
QY 361 GSTKASLQGAIFVLKN---ATQQLNFNDTNVWEGTEANATEYTTGADGIIITIGLKEG 417  
DB 1096 PRTDAD-----KNDPAGKQDVNVGNETPKAE-----DSIGNLPDLPGK 1133  
QY 418 TYLYVE---KKAPLGYNLLDNSQKVL--GDGATDTTNSDNLNVNPTVE----- 461  
DB 1134 TVAFETPVDATPG-----DKPAKVVVVTPDGSKDTVDVTVKVVDPRTDADKNDPAGKQ 1189  
QY 462 --NNKTELPSTGGIGTTIFYIIGAILVIGAGIVLVARR 499  
DB 1190 QVNGKGNKLPATGENATPFNVVALLTIMSSVGLLSVSKKK 1229

RESULT 30  
S28634  
adhesin AIDA-I precursor - Escherichia coli plasmid pIB6  
C;Species: Escherichia coli  
C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 08-Oct-1999  
C;Accession: S28634; S22680; S28881; S72657  
R;Benz, I.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S28634  
A;Accession: S28634  
A;Molecule type: DNA  
A;Residues: 1-1286 <BEN>  
A;Cross-references: EMBL:X65022; NID:g42254; PIDN:CAA46156.1; PID:g42255  
R;Benz, I.; Schmidt, M.A.  
Mol. Microbiol. 6, 1539-1546, 1992  
A;Title: AIDA-I, the adhesin involved in diffuse adherence of the diarrhoeagenic Escheric  
A;Reference number: S22680; MUID:92326638; PMID:1625582  
A;Accession: S22680  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 839-1286 <BE2>  
A;Cross-references: EMBL:X65022  
A;Experimental source: strain 2787  
A;Accession: S28881  
A;Molecule type: protein  
A;Residues: 50-56 <BE3>  
R;Suhr, M.; Benz, I.; Schmidt, M.A.  
Mol. Microbiol. 22, 31-42, 1996  
A;Title: Processing of the AIDA-I precursor: removal of AIDA and evidence for the outer m  
A;Reference number: S72657; MUID:97055419; PMID:8899706  
A;Accession: S72657  
A;Molecule type: protein  
A;Residues: 847-856 <SUH>  
A;Experimental source: DAEC strain 2787  
C;Genetics:  
A;Genome: plasmid pIB6  
C;Keywords: membrane protein

F:1-49/Domain: signal sequence #status predicted <SIG>  
F:50-1286/Product: adhesin AIDA-1 #status predicted <MAT>

Query Match 7.0%; Score 180; DB 2; Length 1286;  
Best Local Similarity 20.6%; Pred. No. 0.046;  
Matches 125; Conservative 76; Mismatches 185; Indels 220; Gaps 30;

Qy	51	VDAEI-DNAVSDSNKDGASYLIPQGEAEYKASTDFNS-----LFTTTTNG	97
Db	164	IFSGGITDSTNIS-----SGQQRVSGGVA---SNTTINSGAQNILSEGAISTHSSG	216
Qy	98	GRYVVTKKDTASANEIATWAKS-----ISANTTPVSTVTSNNDGTEVINVSQYGYVVS	152
Db	217	GNQYI-----SAGANATEIIVNSGFORVNSGAVATGVL---SGGTQ--NVSSGSAIST	267
Qy	153	STVNGAVIMVTSVTPNATIHENKNTDATMGDGGGKTVD-----QKTYSVG	197
Db	268	SVYNSG---VQTVFAGATV---TDITVNSGGNQNISSGGIVSETTVNVSGTQNIYSGG	319
Qy	198	DTVKYITYKNAVNYHGT-----EKVQYVVKDTPWSASV-----VDLNEGSY-EVT	243
Db	320	SALSANIKGSQIVNSEGTAINTLVSDGYQHIRNGGIASCTIVNQSGYVNISSGYAEST	379
Qy	244	ITDGSNITTLTQSEKAT-----GKYNLLENNFTITIPWAATNTPTGNTQNGANDDF	298
Db	380	IINSGGTLRVSDGYARGTILNNSGRENVSNGVSYNAMI-----NT--GGNQYI	427
Qy	299	FYKG-----INTI-----TVTYTGVLKSGAKPGSADLPEN-----	328
Db	428	YSDGEATAAIVNTSGFORINSGGTAPVQNSVVVTRTVSSAAKPFDAEVYSGGQTVYLWR	487
Qy	329	-----TNIAINPNT-----SNDDPGQKVTVRDQI--TIKKIDG-----STK	364
Db	488	GIWYSNFLTAVMSMFPCTASGANVNLSGRLNAPAGNVVGTILNQEGRQVYVSGATATSV	547
Qy	365	ASLQGAIFVLKN--ATCQFLNFNDTNVWEGTEANATEYTTGA-----DGIITITGLKE	416
Db	548	GNNEGREYVLSGGITDGTVLNSGGLQAVSSGGKASATVINEGGAQFVYDGGQVGTGNIKN	607
Qy	417	GTYYLVVEKKAPL-----GYNLL-----DNSQKVLGDG-----	444
Db	608	GGTIRVDSGASALNIALSSGGLFTSTGATLPCLTTMAALSVSQNHASNIVLENGGLLRV	667
Qy	445	-----ATDIT-----NSDNLNVPVENKNG-----T	466
Db	668	TSGGTATDTVNSAGRLRIDGGGTINGTTTINADGIVAGTNIQNDGNFILNLAENYDFET	727
Qy	467	ELPSTG	472
Db	728	ELSGSG	733

Search completed: July 22, 2004, 10:29:19  
Job time : 24 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 09:59:46 ; Search time 13 Seconds  
(without alignments)  
2010.709 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKKIQSLVSLAFGMAV.....GAILVIGAGIVLARRRLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284	11.0	537	1	TEE6_STRPY
2	258	10.0	533	1	FM1_ACTVI
3	231.5	9.0	534	1	FM2_ACTNA
4	185.5	7.2	1025	1	SLAP_CAUCR
5	180	7.0	1286	1	AIDA_ECOLI
6	179.5	7.0	2021	1	OMPA_RICCN
7	178	6.9	1041	1	EGT2_YEAST
8	177.5	6.9	1183	1	CNA_STRAU
9	176	6.8	1276	1	PMP2_CHLPN
10	169	6.6	2249	1	OMPA_RICRI
11	167.5	6.5	939	1	EAE_ECO27
12	167.5	6.5	1902	1	P1P_LACIC
13	167	6.5	2003	1	YDBA_ECOLI
14	165.5	6.4	1902	1	P2P_LACIC
15	165	6.4	1643	1	OMPB_RICPR
16	165	6.4	1655	1	OMPB_RICCN
17	164.5	6.4	1902	1	P3P_LACIC
18	164	6.4	1569	1	YPUA_ECOLI
19	164	6.4	1953	1	BIGA_SALTY
20	162.5	6.3	1902	1	P2P_LACPA
21	159	6.2	2334	1	WAPA_BACSU
22	158	6.1	1004	1	SLPO_BACBR
23	158	6.1	1039	1	AG43_ECOLI
24	158	6.1	1228	1	SLAP_BACST
25	157.5	6.1	670	1	VG50_HSV11
26	157.5	6.1	1656	1	OMPB_RICJA
27	157	6.1	439	1	SLAP_LACHE
28	156.5	6.1	666	1	MUR2_ENTHR
29	156.5	6.1	1020	1	BCA_STRAG
30	156.5	6.1	1861	1	APU_THETU
31	156.5	6.1	2358	1	YEEJ_ECOLI
32	155	6.0	2680	1	YEEJ_ECO57
33	153.5	6.0	497	1	FLIC_ECOLI

## RESULT 1

ID	TEE6_STRPY	STANDARD	PRT	537 AA
AC	P18481			
DT	01-NOV-1990 (Rel. 16, Created)			
DT	01-NOV-1990 (Rel. 16, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Trypsin-resistant surface T6 protein precursor (T6 antigen).			
GN	TEE6			
OS	Streptococcus pyogenes			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			

## ALIGNMENTS

34	153.5	6.0	1300	1	120K_RICRI
35	153.5	6.0	1554	1	OMP_B_RICRI
36	152.5	5.9	934	1	EAE_ECO57
37	152.5	5.9	939	1	SLAP_CAMPE
38	149	5.8	936	1	PMP7_CHLPN
39	148.5	5.8	490	1	TA53_TREDE
40	148.5	5.8	1075	1	FLO5_YEAST
41	148	5.7	738	1	Y013_BPL2
42	148	5.7	1250	1	YFAL_ECOLI
43	148	5.7	1256	1	MRP_STRSU
44	147.5	5.7	465	1	SLAP_LACBR
45	147	5.7	1460	1	PMP_C_HLMU
46	145	5.6	1325	1	YDBK_ECOLI
47	144.5	5.6	863	1	YEO_ECOLI
48	144.5	5.6	1176	1	SLAP_BACSH
49	144	5.6	670	1	YFGG_SCHPO
50	144	5.6	1848	1	CBPA_CLOCL
51	143.5	5.6	1645	1	OMP_RICTY
52	143.5	5.6	1672	1	PMPB_CHLMU
53	143.5	5.6	5703	1	MUSB_HUMAN
54	143	5.5	918	1	YMOB_CAEEL
55	143	5.5	935	1	EAE_ECO11
56	143	5.5	1967	1	CD87_DROME
57	143	5.5	3178	1	YS89_CAEEL
58	142.5	5.5	806	1	SUBV_BACSU
59	142.5	5.5	1754	1	PMPB_CHLTR
60	141.5	5.5	762	1	SLAP_ACEKI
61	141.5	5.5	1065	1	SED4_YEAST
62	141.5	5.5	1289	1	VG34_BPT4
63	141	5.5	2710	1	TOXA_GLODI
64	140.5	5.4	881	1	PRY3_YEAST
65	140.5	5.4	1186	1	CEAA_BACTS
66	140	5.4	313	1	KRE1_YEAST
67	140	5.4	444	1	SLAP_LACAC
68	140	5.4	827	1	CSG_HALVO
69	140	5.4	1628	1	NAGH_CLOPE
70	139.5	5.4	1243	1	VG37_BPK3
71	139.5	5.4	1419	1	ALAI_CANAL
72	139	5.4	1537	1	FLO1_YEAST
73	139	5.4	2432	1	Y43R_IRV6
74	138.5	5.4	1256	1	ATL_STAAU
75	138	5.4	880	1	LYTD_BACSU
76	136.5	5.3	527	1	VG12_BPT4
77	136.5	5.3	1036	1	HP12_DEIRA
78	136.5	5.3	1258	1	ICEN_ERWHE
79	136	5.3	905	1	HXA1_HAEIN
80	136	5.3	1853	1	CIPA_CLOTM
81	136	5.3	2602	1	FLNB_HUMAN
82	135.5	5.3	466	1	FLID_SALTY
83	135	5.2	995	1	YI09_YEAST
84	134.5	5.2	754	1	PLE3_CANAL
85	134	5.2	1778	1	N189_SCHPO
86	133.5	5.2	930	1	PMP8_CHLPN
87	133.5	5.2	1018	1	FNBA_STAAU
88	133.5	5.2	1034	1	ICEN_PANAN
89	133.5	5.2	1150	1	APMU_PIG
90	133.5	5.2	1562	1	SPAP_STRMU

P14914	rickettsia
Q53047	r outer mem
P43261	escherichia
P35827	campylobact
Q92898	chlamydia p
P18164	treponema d
P38894	saccharomyc
P42548	bacterioph
P45508	escherichia
P32653	streptococc
Q05044	lactobacill
Q9PJY1	chlamydia m
P32051	escherichia
P33924	escherichia
P38337	bacillus sp
O13854	schizosacch
P38058	clostridium
P96989	r outer mem
Q9PJY2	chlamydia m
Q9HC84	homo sapien
P34487	caenorhabdi
O31000	escherichia
Q9VG95	drosohila
Q09624	caenorhabdi
P29141	bacillus eu
O84418	chlamydia t
P22558	acetogenium
P25365	saccharomyc
P18771	bacterioph
P16154	clostridium
P47033	saccharomyc
Q45710	bacillus th
P17260	saccharomyc
P35829	lactobacill
P25062	halobacteri
P26831	clostridium
Q38394	bacterioph
O13368	candida alb
P2768	saccharomyc
P18305	chilo iride
P52081	staphylococ
P39848	bacillus eu
P10930	bacterioph
P13126	deinococcus
P16239	erwinia her
P44602	haemophilus
Q06851	clostridium
O75369	homo sapien
P16328	salmonella
P40442	saccharomyc
Q9UVX1	candida alb
Q9UTK4	schizosacch
Q92393	chlamydia p
P14738	staphylococ
Q47879	pantoea ana
P12021	sus scrofa
P23504	streptococc





FT MOD RES 499 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
SQ SEQUENCE 533 AA; 56899 MW; F00299559C702FC4 CRC64;

Query Match  
Best Local Similarity 10.0%; Score 258; DB 1; Length 533;  
Matches 146; Conservative 67; Mismatches 223; Indels 126; Gaps 30;

QY 13 SLAFGMVSPVPIAFAPAEAGTITVQDTQKGYATYKAYKVFDAEIDNANVSDNSKDGASYL 72  
DB 17 TLAAAGALVAP-TGAAPADPNGSTI-DPDAATLTTLVHKCEQDTNGVREGTGNED- 69

QY 73 IFQKREAEKASTDFNSLFT- -TTNGRTTV-TKDTASANEI-ATWAKISANT 124  
DB 70 -PQ--AECKPSVDVEFTIKLNVLDLTVYDGHKTLADLKGDVVKAGALKSTTVQKITCA 125

QY 125 TPVSTVSTESNDGTVEINVSQYGYVSVSTVNGAVI- -MVT--SVTP- -N 169  
DB 126 NGLASFDTAQ- -TEV- -GAYLVSETRPDVKVPAEDFVVLPMINPQDTAKWYN 176

QY 170 ATIHENKNTDATWGDGGKTV-DQKYSVGDVTYKITYK-NAVNHGTEKVVYQVVKDTM 227  
DB 177 VHVYPKNTLS- -GVDKQVTDKPAKSGRDTITYTITSIPKVDYPGGARIKRYEVDRL 232

QY 228 - - - - -PSASVVDLMEGSEVETITDGSNGITTLTQSEKATGKYNLLNENNFIT 277  
DB 233 DKRIKKEALTPVKIV- -GQNEVTLAB-TTDTYLTIT- -AEKG- - - - -DHNWATIQ 277

QY 278 IPWAATNPTGNTGANDDFYKGIN- - -ITVTVTVGLKSGAKPGSADLPENTNIAT 334  
DB 278 LTEEGRRKASEARYNGETKQLVTLNAKFAAVALNGLDLSNTAGLIPNDSP-NFTWDPN 336

QY 335 NPNTSNDG- - - - -QKTVRDQGITTKI- - -DGSTKASLQGAIFVL- - -KNATQGLNF 384  
DB 337 NPGITTDIPGIPPTPVLKSYGVKVLTKGTDLADKTKYNGAQFOVECTKTASGATLRD 396

QY 385 ND- - -TNVWEGTEANATEVTTGADGIITITGLK- - - - -EGTYLVKKAP 427  
DB 397 SPSTQTVPLTIGGEKTEFTAGQGVETINLRANDYNGAKKQDLTDEYVCLVETKAP 456

QY 428 LGYN- - - - -LLDNSQKVLGDGATDTTNSNLLNPTVNNKGTLPSTGGIGTT 477  
DB 457 EGYNLQADPLPRVLAERKAEK- - -KAATEVTVTD- - - - -IPKNAGFRLPLTANGVI 505

QY 478 IFYIIGAILVIGAGIVLVARR 499  
DB 506 FLTIAGALVAGGVAVYANKR 527

RESULT 3  
FM2\_ACTNA STANDARD; PRT; 534 AA.  
AC P12616;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Fimbrial subunit type 2 precursor.  
OS Actinomyces naeslundii.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Actinomycineae; Actinomycetaceae; Actinomycetes.  
OX NCBI\_TaxID=1655;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WWU45;  
RX MEDLINE=8314866; PubMed=2900829;  
RA Yeung M.K., Cisar J.O.;  
RT "Cloning and nucleotide sequence of a gene for Actinomyces naeslundii  
WU45 type 2 fimbriae.";  
RL J. Bacteriol. 170:3803-3809(1988).  
CC -!- FUNCTION: MAJOR FIMBRIAL SUBUNIT OF ACTINOMYCES NAESLUNDII.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
CC -!- SIMILARITY: TO FIMBRIAL SUBUNIT 1 OF ACTINOMYCES VISCOSUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M21976; AAA21931.1; -;  
DR PIR; A32347; A32347.  
DR InterPro; IPR008454; Cna\_B.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF05738; Cna\_B; 2.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR TIGFAMS; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
KW Fimbria; Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 1 32  
FT CHAIN 33 495 FIMBRIAL SUBUNIT TYPE 2.  
FT PROPEP 496 534 REMOVED BY SORTASE (POTENTIAL).  
FT SITE 492 496 LPXTG SORTING SIGNAL (POTENTIAL).  
FT MOD\_RES 495 495 AMIDE-LINKING TO CELL WALL (POTENTIAL).  
SQ SEQUENCE 534 AA; 56574 MW; 6C56C3AB493D5751 CRC64;

Query Match 9.0%; Score 231.5; DB 1; Length 534;  
Best Local Similarity 25.6%; Pred. No. 6.6e-06;  
Matches 141; Conservative 52; Mismatches 202; Indels 155; Gaps 28;

QY 54 AEIDNANVSDNSK- - - - -DGASYLIPQKGEA- - -EYKASTDFNSLFTTTTNG 97  
DB 30 AQAEANAHGDIINTEALGSLTIHKHLNGDNIPGAPDGTASNDGKAPVSGVQFTAYEIN 89

QY 98 GRVYVTKDTSANEIATWA- - -KSISAN- - -TPVSTVSTES- - -NNDGTEVINVS 144  
DB 90 GIDLKTSQWAKVNAITNGAIPDNACANPGQPTLPNTFRSSRVSGDTRDGEAKIESL 149

QY 145 QGYHYVSVSTVNGAVI- - - - -MVTSTVTPNATHEKNTDATW- - - - -GDGSGK 187  
DB 150 PVKALVCEETTPGIVQKAKPFVVTIHPNNTA- - -AKADGTWLYDVHVYPKNEKIEVAK 206

QY 188 TV-DQKT- - -YSVGDITVKY- - -TITYKNAVNVHGTQVYVVKDTMPSASVVDLNEGSY 240  
DB 207 TIEDORNGYIVGSKVFPVSVSTLPKLDNSY- - -KYQP- - -KDT- - -LDNRLLK 253

QY 241 EVTITDGSNGITTLTQSEKATGKYNLLNENNFITIPWAATNPTGNTGANDDFY 300  
DB 254 QVTATDVTLGGLTDLDEGTD- - -YTLGTDGQTVTVTF- - - - -NQNGLSK- - -L 294

QY 301 KGI- - -NTITVTVTVGLK- - -SGAKPGSADLPENTNIATNP- - - - -NTSNDGPGQKVT 348  
DB 295 KGNPGQKLQAVFPGVYSEVGDGSIINNTAQLISDTTYAEQPPAPETPPANPDNPPTTEQVT 354

QY 349 VRDGOITIKKIDGS- - -TKASLQGAIFVL- - - - -KNATGQFLNFDNTNNVSW 392  
DB 355 SKWGLTLTIKKVDGNDRSGDKGLKGAEFQIKAKDAYADTCSPEADGQPLTN- - - - - 407

QY 393 GTEANATEYTTGADGIITITGL- - - - -KEGYVLVVEKAPGLNLLDN 435  
DB 408 - - -GESITFTTGGEGTINFKALFVSDSVQDTRDNRVDAPHRCYVLVETKAPAGVVLPA 463

QY 436 SQKVIL- - -GDGATDTTNSDNLNLYNPTVNNKGTLPSTGGIGTTFIYIG-AILVIGAG 491  
DB 464 ASRAITVFPAGVTVQVVIDN- - -VKQSVPG- - - - -LPLTGANGMLILTASGAALLMIAGV 516

QY 492 IVLVARRRLR 501  
DB 517 SVLVARYRER 526

RESULT 4  
SLAP\_CAUCR  
ID SLAP\_CAUCR STANDARD; PRT; 1025 AA.  
AC P35828; Q46015; Q9RF12;

DT 01-JUN-1994 (Rel. 29, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE S-layer protein (Paracrystalline surface layer protein).  
 GN R8AA OR CC1007.  
 OS Caulobacter crescentus.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;  
 OC Caulobacteraceae; Caulobacter.  
 OX NCBI\_TaxID=155892;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=93007489; PubMed=1393820;  
 RA Gilchrist A., Fisher J.A., Smit J.K.;  
 RT "Nucleotide sequence analysis of the gene encoding the Caulobacter  
 RT crescentus paracrystalline surface layer protein.";  
 RL Can. J. Microbiol. 38:193-202(1992).  
 RN [2]  
 RP REVISIONS TO 376; 636 AND 842-843.  
 RA Awram P.;  
 RN Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=US3001;  
 RA Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;  
 RT "The secretion signal of C. crescentus S-layer protein is located in  
 RT the C-terminal 82 amino acids of the molecule.";  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=21173698; PubMed=11259647;  
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
 RA DeBoy R.J., Dodson R.S., Durkin A.S., Gwinn M.L., Haft D.H.,  
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
 RA Uitterback T., Tran K., Wolf A., Vamathevan J., Esmolaeva M., White O.,  
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
 RT "Complete genome sequence of Caulobacter crescentus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
 RN [5]  
 RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=89008089; PubMed=3049545;  
 RA Fisher J.A., Smit J.K., Agabian N.;  
 RT "Transcriptional analysis of the major surface array gene of  
 RT Caulobacter crescentus.";  
 RL J. Bacteriol. 170:4706-4713(1988).  
 RN [6]  
 RP CHARACTERIZATION.  
 RC STRAIN=ATCC 19089 / CB15;  
 RX MEDLINE=98292737; PubMed=9620954;  
 RA Awram P., Smit J.K.;  
 RT "The Caulobacter crescentus paracrystalline S-layer protein is  
 RT secreted by an ABC transporter (type I) secretion apparatus.";  
 RL J. Bacteriol. 180:3062-3069(1998).  
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY  
 CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A  
 CC PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.  
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-  
 CC LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER  
 CC (TYPE I) SECRETION APPARATUS.  
 CC -1- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE  
 CC SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A  
 CC SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF062345; AAC38665.2; --  
 CC EMBL; AF193063; AAF19365.1; --  
 CC EMBL; AB005779; AAK22991.1; ALT INIT.  
 CC FIR; A48995; A48995.  
 CC HSSP; P22629; LSWC.  
 CC TIGR; CC1007; --  
 CC InterPro; IPR001343; Hemlysn Ca bind.  
 CC Pfam; PF00353; hemolysinbind; 3.  
 CC PRINTS; PR00313; CABNDNGRPT.  
 CC KW Cell wall; S-layer; Calcium-binding; Complete proteome.  
 CC FT INIT MET 0  
 CC SQ SEQUENCE 1025 AA; 98001 MW; AD7A326E136D8AC CRC64;  
 CC -----  
 CC Query Match 7.2%; Score 185.5; DB 1; Length 1025;  
 CC Best Local Similarity 22.1%; Pred. No. 0.0051;  
 CC Matches 131; Conservative 66; Mismatches 240; Indels 157; Gaps 26;  
 CC -----  
 CC QY 6 IQSLIVASLAFGMVSPVTPPIAFAAETGITVQDTQKGATYKAYKVFDAEIDNANVSDSN 65  
 CC DB VQAAVTLPTGTGTISGIE--TNVVTSGAAITLNTSSGVT-----GLTALNTN 343  
 CC QY 66 KDGASYLIPQKKEAEYKASTDFNSLFTTTNGG-----RTYVTK-KDTASANEIATWAK 118  
 CC DB TSGAAQTVTAGAGQNLTAFTAAQAANNVAVDGGANVTVASTGVTSGTTTGVGANSASGTV 403  
 CC QY 119 SIS-ANTTTPVST-----VTESNNDGTEVINVSQY 148  
 CC DB SVSVANSSTTTGAIAVGTGTAIVTAQAGNAVNTLTQADVTVTGNSTTAVTTQTAA 463  
 CC QY 149 YVVSSTVN---NGAVIMVTSVTPNATIHENKNTDAGWGGGKTVDQKTVSGDVTVKYIT 205  
 CC DB ATAGATVAGRVNGAVITIDSAASATACKIATVTLGSGAATIDSSALT----- 513  
 CC QY 206 YKNAVNYHTEK---YQYVIKDTWPSAVVDLNEGSYEVT--ITDGGG---NITTLQ 256  
 CC DB ---TVNLSTGTSGLIGRGALTAT-PTANTLTNLVNGLTITGCAITDSEAAADDGPTTINI 569  
 CC QY 257 GSEKATGKYNLENNFTIIPWATNTPTGNTQNGANDFFYKGINITVT----- 309  
 CC DB AGSTASSTIASLVAADATTINISGDARVITTSHTA-----AALTGITVNSVGATL 620  
 CC QY 310 ----YTGVLKSGAKPGSADLPENTNIA-----TINPNTSNDPQGVTVTRDG-QIT 356  
 CC DB GAELATGLVFTGGAGADSILLGATTKAIVMGAGDDTTVSSATLGAGGSVNGDGTDLV 680  
 CC QY 357 KIDGSTKASLOGAIFVLKNATGQF-----LNFNDTNVNVGTEANATEYTT 403  
 CC DB ANVNGSS-----FSADPAFGPFETLRVAGAAAGSHNANGFTALQLGATAGATTFTN 732  
 CC QY 404 GADGI-ITITGLKEGTYLVEKKAPLG---YNL-LDNSQKVLGD-----GAT 446  
 CC DB VAVNVGLTVLAAPTGTITVTLANA-TGTSVDVNTLSSSALAAGTVALAGVETVNIAT 791  
 CC QY 447 DTTNS---DNLVNPET-----VENNKGTLPSTGGTIGTIFVIICAILIVIGAG 491  
 CC DB DINTTAHVDTLTLOATSAKSIIVTGNAGLINTGNTAVTSP---DASAVTGTG 842  
 CC -----  
 CC RESULT 5  
 CC AIDA ECOLI STANDARD; PRT; 1286 AA.  
 CC ID AIDA ECOLI  
 CC AC Q03155;  
 CC DT 01-JUN-1994 (Rel. 29, Created)  
 CC DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Adhesin aida-I precursor.  
 CC GN AIDA-I.  
 CC OS Escherichia coli.  
 CC OG Plasmid pIB6.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Escherichia.

```

OX NCBI_TaxID=562;
RN
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-56.
RC STRAIN=O126:H27 / 2787;
RX MEDLINE=92326638; PubMed=1625582;
RA Benz I., Schmidt M.A.;
RT "AIDA-I, the adhesin involved in diffuse adherence of the
RT diarrhoeagenic Escherichia coli strain 2787 (O126:H27), is
RT synthesized via a precursor molecule.";
RL Mol. Microbiol. 6:1539-1546(1992).
CC -!- FUNCTION: THIS IS AN ADHESION PROTEIN NECESSARY FOR THE DIFFUSE
CC ADHERENCE OF DIARRHEA-CAUSING ENTEROPATHOGENIC ESCHERICHIA COLI
CC TO EPITHELIAL CELLS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X65022; CA446156.1; --
DR PIR; S28634; S28634.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR004899; Pertactin.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF03212; Pertactin; 1.
DR TIGRfam; TIGR01414; autotrans_bar1; 2. Plasmid.
KW Cell adhesion; Signal; Outer membrane; Adhesin AIDA-I.
FT CHAIN 1 49
FT PROPEP 50 ? ADHESIN AIDA-I.
FT SIGNAL 1 49
SQ SEQUENCE 1286 AA; 132271 MW; B2A00F72AC05FB34 CRC64;

Query Match 7.0%; Score 180; DB 1; Length 1286;
Best Local Similarity 20.6%; Pred. No. 0.013;
Matches 125; Conservative 76; Mismatches 185; Indels 220; Gaps 30;

QY 51 VFDAEI-DNANVSDNKDGASLYIQGKEAYKASTDENS-----LFTITTING 97
DB 164 IFSGGITDSTNLS-----SGQQRVSSGVA-----SMTTNSGAQNILSEGAISTHSSG 216
QY 98 GRTYVTKKDTASANIATWAKS-----ISANTPTVSTTESNNDGTETVINVSQYGYVVS 152
DB 217 GNQYI-----SAGANATEIVNSGGFQRVNSGAVATGTVL---SGGTQ--NVSSGSGAIST 267
QY 153 STVNNGAVIMVTSVTPNATIEHKNTDATWGDGGKKTVD-----QKTYSVG 197
DB 268 SVYNSG---VQTVFAGATV---TDTTVNSGQNISGGIVSETTVNSVSGTQNIYSGG 319
QY 198 DTVKTYITIKYNAVYHGT-----EKYQYVIKDTMPSASV-----VDLNEGSY-EVT 243
DB 320 SALSANIKSQSIVNSGEGTAINTLVSDGGYQHRNGGIASGITVNSGYVINSSSGYABST 379
QY 244 ITDGSNITTLTQSGEKAT-----GKYNLLENNNFITIPWAATNTPTGNTQNGANDDF 298
DB 380 IINSGLTLRLVSDGYARGCTILNNSGRENVSNGVSYNAMI-----NT--GGNQYI 427
QY 299 FYKG-----INTI-----TVYTVGLKSGAKPGSADIAPEN----- 328
DB 428 YSDGEATAIIVNTSGFQRINSCTAPVQNSVVVTRTVSSAAKPFDAEYVSGKQTVLWR 487
QY 329 -----TNIATINPT--SNDPQGVTVRDQI--TIKKIDG-----STK 364
DB 488 GIWYNFLFVMSMPFGTASGANVLSGLNFAFAGVNVGTILNQBGQYVYSGATATSTV 547
QY 365 ASLQGAIFVLKQ--ATGQPLNFNDNNVWEGTEANATEYTCG-----DGIITITGLKE 416
DB 548 GNNEGREYVLSGIGTDTVLSNGGLQAVSSGSKASATVINEGGAQFVVDGGQVGTGNTIKN 607

```

---

```

QY 417 GTYYLVEKKAPL-----GYNLL-----DNSQKVLGDG----- 444
DB 608 GGTIRVDSGASALNIALSSGGLNFTSTGATLPETLTAALSVSONHASNVILENGLLRV 667
QY 445 -----ATDTT-----NSDNLVNPVTVENKKG-----T 466
DB 668 TSGGTATDTTIVNSAGRLRIDGGTGTGTTTINADGIVAGTNIQDNGNFILNLAENVDPET 727
QY 467 ELPSTG 472
DB 728 ELSSGS 733

RESULT 6
OMPA_RICCN
ID OMPA_RICCN STANDARD; PRT: 2021 AA.
AC Q52657; P95591; P95592; P95594; Q52667; Q52668; Q52669;
AC Q52670; Q52674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (Cell surface
DE antigen) (rOmpA) (rOmp A).
GN OMPA OR RCL273.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Croquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
RT conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
RT analysis of restriction fragment length polymorphism of PCR-amplified
RT DNA of the gene encoding the protein rOmpA.";
RL J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.E., Roux V.;
RT "Phylogenetic analysis of spotted fever group rickettsiae by study
RT of the outer surface protein rOmpA.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-
CC layer with hexagonal symmetry.
CC -!- PTM: Glycosylated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```





CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----

DR EMBL; M81736; AAA20874.1; --  
DR PDB; 1AMX; 24-JUN-98.  
DR PDB; 1D20; 27-SEP-00.  
DR PDB; 1D2P; 27-SEP-00.  
DR InterPro; IPR008966; Adhes bact.  
DR InterPro; IPR008454; Cna\_B  
DR InterPro; IPR008970; Cna\_B unit.  
DR InterPro; IPR008456; Collagen bind.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF05738; Cna\_B; 7.  
DR Pfam; PF05737; Collagen bind; 1.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; Repeat; Signal; 3D-structure.  
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.  
FT SIGNAL 1 29  
FT CHAIN 30 1154  
FT PROPEP 1155 1183  
FT DOMAIN 151 318  
FT DOMAIN 533 1093  
FT DOMAIN 1093 1157  
FT REPEAT 533 719  
FT REPEAT 720 906  
FT REPEAT 907 1093  
FT SITE 1151 1155  
FT MOD RES 1154 1154  
FT STRAND 174 179  
FT TURN 182 183  
FT TURN 185 186  
FT STRAND 187 194  
FT TURN 196 197  
FT STRAND 201 201  
FT STRAND 205 211  
FT STRAND 215 228  
FT TURN 229 230  
FT STRAND 232 234  
FT STRAND 239 246  
FT TURN 248 249  
FT STRAND 251 255  
FT TURN 256 259  
FT STRAND 260 265  
FT STRAND 267 270  
FT TURN 271 272  
FT STRAND 273 283  
FT TURN 286 287  
FT STRAND 290 299  
FT STRAND 301 301  
FT TURN 302 303  
FT STRAND 307 311  
FT STRAND 314 317  
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072B57D76 CRC64;

Query Match 6.9%; Score 177.5; DB 1; Length 1183;  
Best Local Similarity 21.8%; Pred. No. 0.017;  
Matches 152; Conservative 65; Mismatches 236; Indels 245; Gaps 35;  
QY 24 TPIAFAETGTTVDQTK--GATYKAYKVFDAEIDN-----AN----- 60  
DB 512 TPEGTKKENGVLVNTKEPTETTSISGEKWD-DKQDQGRPEKVSVLLANGEKVK 570  
QY 61 ---VSDSN-----KDGASYLPQKGEAYKASTFNSLFTTTTNGGRYTVKKDTASANE 112  
DB 571 LDVTSETNWKYEFKDLPKY--DEGKKIETVTEHDVKDYTTDING--TTIINKYTPGETS 626  
QY 113 IATWAKSISANTT-----PVSVTVTESNNDG-----TEVINVS-----Q 145

DB 627 -ATVTKNWDNNQDGRKPTTEIKVELYQDGRKATKTAALNESNNWTHWTGLDEKAKGQQ 685  
QY 146 XGYI-----VYSVTVN--GAVIMVTSVTNATHEKNTDATWGD---GGGKTVDQ 191  
DB 686 VKYTVBELTKVGTYTHVDNDMGNLIVNKYTPETT--SISGEKVWDDKDNQDGRPEK 743  
QY 192 -----KTVSGDVTVKYTYTYKNAVNYHGTEKVOYVVKDTMPSASVDLN--- 236  
DB 744 VSVNLLADGEKVKTLDTSETNWKYEFKDLPKYDEGKKI-EYTVTEHDVKDYTTDINGTT 802  
QY 237 -----EGSYEVTTT--DGSNG-----ITLTQSEKATGKYNLENNNFTTIP 279  
DB 803 IINKYTPGETSATVTKNWDNNQDGRKPTTEIKVELYQDGRKATKTAALNESNNWTH-- 860  
QY 280 WAAVTNPT-----GNTQNGANDFFYKGINTTITVYT----- 311  
DB 861 WTGLDEKAKGQOVKYTVBELTKVGTYTHVDNDM--GNLIVNKYTPETTSISGEKVW 917  
QY 312 -----GVKSGAKPGSADLPENTNIATINPTNSDDPGQV--TVRDGQ 353  
DB 918 DDKDNQDGRPEKVSVLLANGEKVKTLDTSETNWKYEFKDLPKYDEGKKIETVTEHD 977  
QY 354 -----ITIKIDGSTKASLQGAIFVLKN-----A 377  
DB 978 VKDYTTDINGTTINKYTPGETSAT-----VTKNWDNNQDGRKPTTEIKVELYQDGA 1031  
QY 378 TQQLFNFDNTNN-----VEWGTENATEYTTGADG-----IT- 410  
DB 1032 TCKTALNESNNWTHWTGLDEKAKGQOVKYTVDELTKVNG--YTHVDNDMGNLIVTN 1089  
QY 411 -ITGLKEGTYLYVEK---KAPLGYNLLDSQKVLGDGATDTNSDNLVNPVTVENKGT 466  
DB 1090 KYTPKPKNPPIYPEKPKDKTPTPKDHSNKVKPTPPDKPSKVDKDDQPKDNKTKPENPLK 1149  
QY 467 ELPSTGG--IGTTFIVIIIGAILVIGAGIVLVARRRLRS 502  
DB 1150 ELPKTKMKIITSWITWVIGIL-----GLYLILRKFNFS 1183  
RESULT 9  
PMP6\_CHLPN STANDARD; PRT; 1276 AA.  
AC Q92899; Q9JRW2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable outer membrane protein pmp6 precursor (Polymorphic membrane protein 6).  
GN PMP6 OR CPN0444 OR CP0309 OR CPB0460.  
OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia. phila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
RL Nat. Genet. 21:385-389 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406 (2000).  
RN [3]



```
FT REPEAT      506 577 E (TYPE II).
FT REPEAT      578 652 F (TYPE II).
FT REPEAT      653 724 G (TYPE II).
FT REPEAT      725 799 H (TYPE II).
FT REPEAT      800 874 I (TYPE II).
FT REPEAT      875 949 J (TYPE II).
FT REPEAT      950 1021 K (TYPE II).
FT REPEAT     1022 1093 L (TYPE II).
FT REPEAT     1094 1165 M (TYPE II).
FT REPEAT     1166 1180 TYPE I (INCOMPLETE).
SQ SEQUENCE    2249 AA; 224333 MW; A9D6646C089DF087 CRC64;

Query Match
Best Local Similarity 6.6%; Score 169; DB 1; Length 2249;
Matches 140; Conservative 53; Mismatches 190; Indels 164; Gaps 32;

QY 12 ASLAFGMVSPVTPPIAFPAETGTTTQDTQKATYKAYKVFDAIDNA-----NVPSNKKD 67
D 12 ASLAFGMVSPVTPPIAFPAETGTTTQDTQKATYKAYKVFDAIDNA-----NVPSNKKD 67
D 741 ATLGGAIVKATTTKLTNAASVLT-----NANAVLTGAIDNTTGGDNVGVNLN 790
QY 68 GASVLIPQGEAEYKASTDF---NSLET-----TTTNGRTVYTKDTASANEIATWAK 118
D 68 GASVLIPQGEAEYKASTDF---NSLET-----TTTNGRTVYTKDTASANEIATWAK 118
D 791 GAL-----SQVTGDIQNTSLATISVGAGTATLGA--VTKATTTKLTNAASVLT 838
QY 119 SLSAN---TTPVSTVTESNDGTEVINVSQVGYVYSSTVWNGAVIMVTSVTPN---AT 171
D 119 SLSAN---TTPVSTVTESNDGTEVINVSQVGYVYSSTVWNGAVIMVTSVTPN---AT 171
D 839 LTANAVLTGAVDNTTGGDNV--VLNL-----NGALSQVTGDIQNTSLAT 893
QY 172 IHEKNTDATWGGGGKTVDDQKYSVGTVKYTIYKNAV-----NVHGTKEYVQYVIKD 225
D 172 IHEKNTDATWGGGGKTVDDQKYSVGTVKYTIYKNAV-----NVHGTKEYVQYVIKD 225
D 884 ISVGAGTATLGGAVIKATTTKLTNAASVLT---TLTNAVALTGAIDNTTGGDNV----- 934
QY 226 TSPASVVDINEGSEYVITDGSN--ITLTQSEKAT-----CKYNLLBENNFTI 276
D 226 TSPASVVDINEGSEYVITDGSN--ITLTQSEKAT-----CKYNLLBENNFTI 276
D 935 -----GVNLNLGALSQVTGDIQNTSLATISVGAGTATLGAIVKATTTKLTNAASVLT 989
QY 277 TIPWAAINT--PTGNTQNGANDDFYKGINITITVYTVGLKSGAKPGSADLPENTN--IAT 333
D 277 TIPWAAINT--PTGNTQNGANDDFYKGINITITVYTVGLKSGAKPGSADLPENTN--IAT 333
D 990 TNPVVVTGAIDNTGNANG-----IVTFTG---NSIVTGVNG--NTNALAT 1030
QY 334 INPNTSNDPQGVTVRDGQITIKIDGSKYAS-----LQGAIFVLKNAATGQFLNF 384
D 334 INPNTSNDPQGVTVRDGQITIKIDGSKYAS-----LQGAIFVLKNAATGQFLNF 384
D 1031 VNVGAG-----LLQVGGVVKANTINLTDNASAVTFTNPVVVTGAIDNTGNANGIVTF 1084
QY 385 NDTNNVWEGTEANATEYTT---GADGLIITG--LKEGTYLVKKAPLGVNLLDNSQV 439
D 385 NDTNNVWEGTEANATEYTT---GADGLIITG--LKEGTYLVKKAPLGVNLLDNSQV 439
D 1085 TGNSTVT--GNVGNALATVNVGA--GLQVGGVVKANT-----INLTDNASAV 1131
QY 440 -----ILGDGATDTT--NSDNLV-----NPTVENNKGTELPSTGGIGTTFIYIIGAILVIG 489
D 440 -----ILGDGATDTT--NSDNLV-----NPTVENNKGTELPSTGGIGTTFIYIIGAILVIG 489
D 1132 TFTNPVVVTGAIDNTGNANGIVTFTGNSIV-----TGDIQNTNAL---ATVNVG 1178
QY 490 AGIVLVA 496
D 490 AGIVLVA 496
D 1179 AGITLQA 1185

RESULT 11
EAE_ECO27
ID -EAE ECO27 STANDARD; PRT; 939 AA.
AC P19809;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Intimin (Attaching and effacing protein) (Eae protein).
GN EAE OR EAEA.
OS Escherichia coli O127:H6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_taxID=168807;
RN [1]
RC SEQUENCE FROM N.A.
STRAIN=O127:H6 / E2348/69;
```

```
RX MEDLINE=91045893; PubMed=2172966;
RA Jerse A.E., Yu J., Tall B.D., Kaper J.B.;
RT "A genetic locus of enteropathogenic Escherichia coli necessary for
RL the production of attaching and effacing lesions on tissue culture
RL cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7839-7843(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O127:H6 / E2348/69;
RX MEDLINE=98254123; PubMed=593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donnenberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
RL from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).
CC -!- FUNCTION: NECESSARY FOR THE PRODUCTION OF ATTACHING AND EFFACING
CC LESIONS ON TISSUE CULTURE CELLS. BELIEVED TO MEDIATE ADHERENCE.
CC -!- SUBCELLULAR LOCATION: Outer surface.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- SIMILARITY: Contains 1 LysM repeat.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; MS8154; AA62775.1; -.
CC EMBL; AF022236; AAC38392.1; -.
CC PIR; I41197; I41197.
CC PDB; 1ESU; 03-OCT-00.
CC PDB; 1F00; 12-JUL-00.
CC PDB; 1F02; 12-JUL-00.
CC InterPro; IPR003344; Big_1.
CC InterPro; IPR003343; Big_2.
CC InterPro; IPR008964; Invasin_intimin.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR002482; LysM.
CC Pfam; PF02369; Big_1; 2.
CC Pfam; PF02368; Big_2; 1.
CC Pfam; PF00059; lectin_c; 1.
CC Pfam; PF01476; LysM; 1.
CC PRINTS; PR01369; INTIMIN.
CC SMART; SM00634; BID_1; 2.
CC SMART; SM00635; BID_2; 1.
CC SMART; SM00257; LysM; 1.
CC Outer membrane; Virulence; 3D-structure.
FT REPEAT 65 113 LYSM.
SQ SEQUENCE 939 AA; 102410 MW; 783C53EB0322DE4D CRC64;

Query Match 6.5%; Score 167.5; DB 1; Length 939;
Best Local Similarity 24.1%; Pred. No. 0.044;
Matches 116; Conservative 42; Mismatches 195; Indels 129; Gaps 23;

QY 27 AFAAETGTTTVDQTKGATYKA-----YKVEDAIDNANVSDSNKDGASVLP 74
D 27 AFAAETGTTTVDQTKGATYKA-----YKVEDAIDNANVSDSNKDGASVLP 74
D 491 ALRSQGGQIQHSQSAQDYQAILPAYVGGSNVYKVTARAYDRNGNSNNVLTITVLS 550
QY 75 QGKEAEYKASTDFNSLFTTTTNGRTYVTKDTASANEIATWAKSISANT---TPVSTVT 131
D 75 QGKEAEYKASTDFNSLFTTTTNGRTYVTKDTASANEIATWAKSISANT---TPVSTVT 131
D 551 NGQVDPQVGVTDFTADTKSAKADGTEAITVTATVKNGVAQANVPVSFNVISGTAVLSAN 610
QY 132 ESNDG-----TEVINVSQYGYVYS-----STVNGAVIMVTSVTPNATHEKNTDAT 180
D 132 ESNDG-----TEVINVSQYGYVYS-----STVNGAVIMVTSVTPNATHEKNTDAT 180
D 611 SANTNGSGKATVTLKSKDFQVNVSAKTAEMTSALANAVIFVDQT--KASITEIKADKT 668
QY 181 WGDGGKKTVDQKYSV---GD-----TVKYTIT---YKNAVNVHGTKEYVQYVIKDTM 227
D 181 WGDGGKKTVDQKYSV---GD-----TVKYTIT---YKNAVNVHGTKEYVQYVIKDTM 227
D 669 TAVANGQ--DAITYTVKVMKGPVSNQOEVTFTTLGLKLSNTEKTDITNGYAKVLTSTT 726
```



QY 228 PSASVVDLN-----EGSYEVITDGSNITTLTGSEKAT-----GKNLLE 269  
 Db 727 PGKLSARVSDVADVKAPEVEFFTTLDIGNIBIVGTGVKGLPTVWLQYGVNLKA 786  
 QY 270 ENNNFTITPWAATNPTGNTQNGAND-DFFYKGINTI-----TVTYTGVLSKARP 320  
 Db 787 SGGNGKVT--WRSANPAIASVDASSQVTLKKGTTTISVISDNOQTAYT----- 835  
 QY 321 GSADLPENTNIATIN-----PNTSNDPDGQKVTVRDGOITIKKIDGSKASLOQAIFVLKN 376  
 Db 836 -----IATPNSLIVPNMS-----KRVTYNDVAVNTCKNPGGKLPSSQ----- 871  
 QY 377 ATGQFLNFNDTNNV--EMGTEANATYTTGADGII-----TITGLKEG---TYVLVEKA 426  
 Db 872 -----NELENVFRWAGA-ANKYBYKSSQTIISVWQTAQDAKSGVASTVDLV-KQN 921  
 QY 427 PL 428  
 Db 922 PL 923

RESULT 12  
 ID PIP\_LACLC STANDARD; PRT; 1902 AA.  
 AC P16271;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PI-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine  
 DE proteinase).  
 GN PRTP.  
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
 OG Plasmid pW05.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1359;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WG2;  
 RX MEDLINE=88149035; PubMed=3278687;  
 RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;  
 RT "Nucleotide sequence of the cell wall proteinase gene of  
 RT Streptococcus cremoris Wg2";  
 RL Appl. Environ. Microbiol. 54:231-238 (1988).  
 CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
 CC GROWTH OF THE BACTERIA ON MILK.  
 CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
 CC specificity, although some substrate preference have been noted,  
 CC e.g. large hydrophobic residues in the P1 and P4 positions, and  
 CC Pro in the P2 position. Best known for its action on caseins,  
 CC although it has been shown to hydrolyze hemoglobin and oxidized  
 CC insulin B-chain.  
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
 CC an amide bond (Potential).  
 CC -!- SIMILARITY: Belongs to peptidase family S8.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M24767; AAA17677.1; -;  
 CC HSSP; P00782; 1S01.  
 CC MEROPS; S08.019; -;  
 CC InterPro; IPR001899; Gram\_pos\_anchor.  
 CC InterPro; IPR003137; PA.  
 CC InterPro; IPR000209; Peptidase\_S8.  
 CC Pfam; PF00746; Gram\_pos\_anchor; 1.  
 CC Pfam; PF02225; PA; 1  
 CC Pfam; PF00082; Peptidase\_S8; 1.  
 CC PRINTS; PR00723; SUBTILISIN.

TIGRFBMS; TIGR01167; LPXTG\_anchor; 1.  
 PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 PROSITE; PS00138; SUBTILASE\_SER; 1.  
 KW Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;  
 KW Signal; Plasmid.  
 FT SIGNAL 1 33  
 FT PROPEP 34 187  
 FT CHAIN 188 1870  
 FT PROPEP 1871 1902  
 FT ACT\_SITE 217 217  
 FT ACT\_SITE 281 281  
 FT ACT\_SITE 620 620  
 FT SITE 1867 1871  
 FT MOD\_RES 1870 1870  
 SQ SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5DOB CRC64;  
 Query Match 6.5%; Score 167.5; DB 1; Length 1902;  
 Best Local Similarity 21.6%; Pred. No. 0.11;  
 Matches 114; Conservative 71; Mismatches 171; Indels 173; Gaps 26;  
 QY 25 PIAFAAETGTTTVDQTKGATYKAVKVFDAEIDNANVSNSKDGASYLIPQCKEAYKAS 84  
 Db 704 PSTVAENGYPVELKDFSTDTKFKL-----TFTNSTTHELTY-----QMDSN 747  
 QY 85 TDFNSLFTTTT--NGRTYVTKKDTASANEIATWAKSISANTTPVSTVTE----- 132  
 Db 748 TDTNAVYTSATDPNSGVLYDKKIDGAA-----IKAGSNITVPAGKTAIEFTLSLPKS 800  
 QY 133 -----SNNDGTEVINVSQYGY-----YVSTVNGAVIV 163  
 Db 801 FQOQOFVEGFLNFKSGDGR-LNLPYMGFFGDMNDGKIVDSLNGITYSPAGGNFGTVPL 859  
 QY 164 TGVTPNATIEKNTDATWGDG-----GGTKVDOKTYSVGDVTKVTTTVKNAVNYHGTEK 217  
 Db 860 TN-----KVTGYGYGVMVDADGNQTVDDQAIASFSSD-----KNAL-YNDISM 902  
 QY 218 VYQYVVKOTMPSASVVVDLNEGSYEYVITDGSN-ITTLTGSEKATGKYNLLENNFTI 276  
 Db 903 KY-YLLR-----NISNVQVDILDGQGNKVTTLSSSTNLTKTYNAHSQQYIYVN 950  
 QY 277 TTPWAAT--NTPTGNTQNGANDDPYKGINTIYTVTVGLKSGAKPGSADLPENTNIATI 334  
 Db 951 APAMDGTYYDQDRGNIKTADDSYTYR-----ISGVPEGGDKRQVDFVPFKLD--- 998  
 QY 335 NENTSNDPDGQKVTVRDGOITIKKIDGST-----KASLOG--AIFVLKQATGQFLNF 384  
 Db 999 -----SKAPTVRHVALSAKTENGKTQYVLTAEAKDGLSLDGTKSVKTA-----I 1043  
 QY 385 NDTNNVEMGTEANATEYTTGADGIITI-TGLKEGTYYLVKRAPGLYNLLDNSQKVLGD 443  
 Db 1044 NEVTNL---DATFTDAGTTADGYTKIETPLSD-----EQAQALGNG--DNSAELVLT 1091  
 QY 444 GATDTTNSDLLVNPVTENNKTELPSTGGIGTTFYIIGAILVTCAGI 492  
 Db 1092 NASNATDQASQVKP-----GSTSPDLI-----VNGGGI 1120

RESULT 13  
 YDBA\_ECOLI STANDARD; PRT; 2003 AA.  
 AC P33666; P76088; P76856; P76857; P76859;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein ydbA.  
 GN YDBA OR B1401/B1405.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RR Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of *Escherichia coli* K-12.";  
RL Science 277:1453-1474 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251357; PubMed=9097039;  
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
RR Kawai H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa M.,  
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,  
RA Sampei G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;  
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome  
corresponding to the 28.0-40.1 min region on the linkage map.";  
RL DNA Res. 3:363-377 (1996).  
RN [3]  
RP SEQUENCE OF 464-2003 FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92190338; PubMed=1665988;  
RA Moszer I., Glaser P., Danchin A.;  
RR "Multiple IS insertion sequences near the replication terminus in  
*Escherichia coli* K-12.";  
RL Biochimie 73:1361-1374 (1991).  
CC -!- SIMILARITY: TO S.TYPHIMURIUM ORF NEAR CYSG (AC P25928).  
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR  
THIS PROTEIN IS INTERRUPTED BY A HYBRID IS2D/IS30C ELEMENT  
BETWEEN AMINO ACIDS 839 AND 840.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000237; AAC74483.1; ALT SEQ.  
DR EMBL; AE000237; AAC74487.1; ALT SEQ.  
DR EMBL; D90778; BAA15009.1; ALT SEQ.  
DR EMBL; D90778; BAA18880.1; ALT SEQ.  
DR EMBL; D90779; BAA18881.1; ALT SEQ.  
DR EMBL; X62680; -; NOT ANNOTATED CDS.  
DR EcoGene; EG11307; ydha.  
KW Hypothetical protein; Complete proteome.  
FT CONFLICT 489 489 I -> V (IN REF. 2).  
FT CONFLICT 495 495 I -> V (IN REF. 2).  
SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220EE CRC64;  
Query Match 6.5%; Score 167; DB 1; Length 2003;  
Best Local Similarity 22.3%; Pred. No. 0.12;  
Matches 136; Conservative 67; Mismatches 196; Indels 210; Gaps 29;  
QY 33 GTTITVDTQKATYKAYKVFDAEIDANVSDNSKDGASVLPQCKEAYKASTDFNSLFT 92  
DB 279 GTMTVDPES-----MGFIQID-GDKAVNNEGSTINGG-----TGTQINGDDA 322  
QY 93 TTITNGRTVTKKDTASANEIATWAKSISANTTPVSTVTSNNDGTEVI-----NVSYG 147  
DB 323 TANNNGKITVDGKD-----STGTEINGNNGKVIQDGLDVSQGG 361  
QY 148 YYY-----VSSVNNCAVIMVTSVT-----PNATHEKNTDATWGGGKTVQDKTYS 195  
DB 362 HGIDITGSDVNDKGTMTVTDPEISIGIQVGDQAVNNEGESAITNGTGTQINGDDAT 421  
QY 196 VGDVTKYITTKYKAVNNGHTE-----KYYQ-----YVTKDTMPSASV----- 232

Db 422 ANNNGKTTVDGKOST---GTEIAGNNGKVIQDGLDVSQGGHGIDITGDSATVDNKGTMT 478  
QY 233 -----VDLNEGSYEYVITDG-----SGNITTLTQOSE----- 259  
Db 479 VTDPEISIGIQDGOAIVNNEG--ESTITNGTGTQINGNDATANNNGKTTVDGKDSGT 536  
QY 260 KATGYNLLLENNPTIT-IPWAATNPTGTQNGANDDFPKYKGINITVITYT---GVLK 315  
Db 537 KIAGNIGIVNLDGSLTGTGAGHVENIGDGTGNVANKG-----IIVSDTSGISGVL 587  
QY 316 SCAPKGSADLPENTNIA-TINPNTSNDPQGVKTVRDGOITI---KKIDGSTRKASLQ- 369  
Db 588 NG-----EGATVNTGDNVNSNEATGFSITNSKVSGLASGMQVDFSTGVDLNGN 638  
QY 370 -----AIFVLK-----NATQFLNFDNTNV-----EWGTEANATEY-----T 402  
Db 639 NNSVTLAAKDLKVVQKATGIVNSGDANTVITGNVLVDKDKTADNAAEYFPDPSGVIN 698  
QY 403 TGADGIITGLKEGYLYLVEKKAPLGYNLLDNSOK-----VILGDGATDTTNSDLNV- 456  
Db 699 YGSDNNVTLDG--KLTVVSDSEVTSRQSNLFDGSAEKTSGLVVIGDGNVNMNGGLELIG 756  
QY 457 --NPTVENNKGTLPSTGGTGTTFYIIG-----AI 485  
Db 757 EKNALADGSQVTSL-RTGYSYTSVIVSGESSVYLVNGDTTISGEFPLGFAGVIRVQDKAL 815  
QY 486 LVIGAGIVL 494  
Db 816 LBIGSGATL 824  
RESULT 14  
P2P LACLC  
ID P2P LACLC STANDARD; PRT; 1902 AA.  
AC P15293;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE PII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-  
associated serine proteinase) (LPI511).  
GN PRT.  
OS *Lactococcus lactis* (subsp. cremoris) (*Streptococcus cremoris*).  
OG Plasmid pLP763  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; *Lactococcus*.  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCDO 763;  
RX MEDLINE=89313288; PubMed=2501630;  
RA Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;  
RT "Molecular characterization of a cell wall-associated proteinase gene  
from *Streptococcus lactis* NCDO763.";  
RL Mol. Microbiol. 3:359-369 (1989).  
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
GROWTH OF THE BACTERIA ON MILK.  
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
specificity, although some substrate preference have been noted,  
e.g. large hydrophobic residues in the P1 and P4 positions, and  
pro in the P2 position. Best known for its action on caseins,  
although it has been shown to hydrolyze hemoglobin and oxidized  
insulin B-chain.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
CC -!- SIMILARITY: Belongs to peptidase family S8.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

Q53020; Q9ZC00;  
30-MAY-2000 (Rel. 39, Created)  
30-MAY-2000 (Rel. 39, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Outer membrane protein B precursor (168 kDa surface-layer protein) (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOmpB) (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
OMP OR SPAP OR SPA OR RP704.  
Rickettsia prowazekii.  
Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
Rickettsiaceae; Rickettsiae; Rickettsia.  
NCBI\_TaxID=782;  
[1]  
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
STRAIN=Breil;  
MEDLINE=91045972; PubMed=2122457;  
Carl M., Dobson M.E., Ching W.M., Dasch G.A.;  
"Characterization of the gene encoding the protective paracrystalline-surface-layer protein of Rickettsia prowazekii: presence of a truncated identical homolog in Rickettsia typhi.";  
Proc. Natl. Acad. Sci. U.S.A. 87:8237-8241(1990).  
[2]  
SEQUENCE FROM N.A.  
STRAIN=Breil;  
Moron C.G., Yu X.J., Walker D.H.;  
"Sequence analysis of ompB of Rickettsia prowazekii.";  
Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE FROM N.A.  
STRAIN=Madrid E;  
MEDLINE=99039499; PubMed=9823893;  
Andersson S.G.E., Zomorodipour A., Andersson J.O.,  
Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
Eriksson A.-S., Winkler H.H., Kurland C.G.;  
"The genome sequence of Rickettsia prowazekii and the origin of mitochondria";  
Nature 396:133-140(1998).  
[4]  
PARTIAL SEQUENCE.  
STRAIN=Breil;  
MEDLINE=92114896; PubMed=1370573;  
Ching W.M., Carl M., Dasch G.A.;  
"Mapping of monoclonal antibody binding sites on CNBr fragments of the S-layer protein antigens of Rickettsia typhi and Rickettsia prowazekii.";  
Mol. Immunol. 29:95-105(1992).  
[5]  
CLEAVAGE SITE.  
MEDLINE=92104668; PubMed=1729180;  
Hackstadt T., Messer R., Cieplak W. Jr., Peacock M.G.;  
"Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein of rickettsiae: identification of an avirulent mutant deficient in processing.";  
Infect. Immun. 60:159-165(1992).  
CC -1- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.  
CC -1- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR.  
CC -1- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.  
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC ENBL; M37647; AAA26390.1; ALT INIT.  
CC EMBL; AF161079; AAD42234.1; -  
CC



Query Match 6.4%; Score 165; DB 1; Length 1655;  
Best Local Similarity 22.8%; Pred. No. 0.12;  
Matches 134; Conservative 67; Mismatches 209; Indels 182; Gaps 29;

QY 30 AETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSN-----KDGA-----SYLP 74  
DB 606 AQTLLNGKIGTVGANKKILQGNISSTKVLSDGVNINELVIGNNGAVQFAHNTYLLT 665  
QY 75 QKKEAEYKASTDPN-----SLFTTTTNGGRT-----YVTKKOTASANEIATWAKSIS 121  
DB 666 RTTNAAGQGGKIIFNPVNNNTTLATGNLGSATNPLAEINFGSKGAANDVTVLNVGKGVN 725  
QY 122 ANTPPVSTVTS-----NNDGTEVINVS-----QYGYVYSSTVNNGAVIMVTSVTPNAT 172  
DB 726 LYATNITTDANVGSFIFNAGGTVISVTVGGQGNKFNFTVALDNGNTTVKFLG---NATP 782  
QY 173 HEKNT---DATWGDGGKTVQDKTYSVG-----DTVKYTIYK----- 207  
DB 783 NGNTTIAANSTLQIGNYTADFVASADGTGIVEFVNTGPIVTLNKQAAPVNAVKQITVS 842  
QY 208 -----NAVNYHGTGKYYQYVYIKDTPSPASVVDLNEGSYEVTITDGSNIITL-- 254  
DB 843 GPGNVVINEIGNAGNYHGA--VTDITAFENSSLGAVVFLPRG---IPFNDAGNRIPLTIK 897  
QY 255 -TQSEKATGKYNLLENNFTIIPWA-----ATNTPGNTQN-----GANDDF 298  
DB 898 STVGKNTATG-----FDVPSVIVLGVDSVIADQGVIGDQNNIVGLGLGSDNDI 945  
QY 299 -----FYKGINTI-----TVTVTG-----VLKSGAKPGSADLPENT-----NIAT 333  
DB 946 IVNATLYAGITINNQQGTVLUSGIPNTPGTGYGLTGIGASKPKQVTFDYNLGN 1005  
QY 334 INPNTSNDPFGQKTVYR-----DQGITIKID-----GSTKASL 367  
DB 1006 LIATNATINDVTVTTGGIAGFDGCKITLGSVNGNVRVFDGILSHSTSMIGTTKAN- 1064  
QY 368 QGAIFVLKNAQTGFLNPDNNVWGTENANETTYTGADGITITGLKGGTYLVLEKAP 427  
DB 1065 NGTVTVYLGNAFVGNIGSDT-----PVASVRFVTSVDSG---GAGLOGNIYSQV---ID 1110  
QY 428 LG-YNLLDSNQKVLGDGATDITNSDNLVNPVNNKTELPSTGIGITTI 478  
DB 1111 FGYNLGSNSVILGGGTGTAINGKINLRN-TLTFASGT---STWGNNTSI 1158

RESULT 17  
P3P\_LACLC STANDARD; PRT; 1902 AA.  
AC F15292;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE PIII-type proteinase precursor (EC 3.4.21.96) (Lactococpin) (Cell wall-  
DE associated serine proteinase).  
GN PRTP.  
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).  
OG Plasmid.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1359;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.  
RC STRAIN=SK11;  
RX MEDLINE=89340435; PubMed=2760036;  
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;  
RT "Primary structure and organization of the gene for a procaryotic,  
RT cell envelope-located serine proteinase."  
RL J. Biol. Chem. 264:13579-13585(1989).  
CC -I- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE  
CC GROWTH OF THE BACTERIA ON MILK.  
CC -I- CATALYTIC ACTIVITY: Endopeptidase activity with very broad  
CC specificity, although some subsite preference have been noted,  
CC e.g. large hydrophobic residues in the P1 and P4 positions, and  
CC Pro in the P2 position. Best known for its action on caseins,  
CC

although it has been shown to hydrolyze hemoglobin and oxidized  
insulin B-chain.  
-I- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
-I- SIMILARITY: Belongs to peptidase family S8.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
EMBL; J04962; AAA03533.1; ALT\_SEQ.  
HSSP; P00782; 2SBT.  
MEROPS: S08.019; -  
InterPro: IPR001899; Gram\_pos\_anchor.  
InterPro: IPR003337; PA.  
InterPro: IPR002029; Peptidase S8.  
Pfam: PF00746; Gram\_pos\_anchor; 1.  
Pfam: PF02225; PA; 1.  
Pfam: PF00082; Peptidase S8; 1.  
PRINTS: PR00723; SUBTILISIN.  
TIGRfam: TIGR01167; LPXKG\_anchor; 1.  
PROSITE: PS00847; GRAM\_POS\_ANCHORING; 1.  
PROSITE: PS00136; SUBTILASE ASP; 1.  
PROSITE: PS00137; SUBTILASE HIS; 1.  
PROSITE: PS00138; SUBTILASE\_SER; 1.  
Hydrolase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;  
Signal; Plasmid.  
FT SIGNAL 1 33  
FT PROPEP 34 187  
FT CHAIN 188 1870  
FT PROPEP 1871 1902  
FT ACT\_SITE 217 217  
FT ACT\_SITE 281 281  
FT ACT\_SITE 620 620  
FT SITE 1867 1871  
FT MOD\_RES 1870 1870  
FT SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;  
Query Match 6.4%; Score 164.5; DB 1; Length 1902;  
Best Local Similarity 20.9%; Pred. No. 0.16;  
Matches 108; Conservative 76; Mismatches 184; Indels 149; Gaps 25;  
QY 25 PIATFAETGTTITVQDTQKGATYKAYKVFDAEIDNANVSDSNKDGASLYLIPQCKEAEYKAS 84  
DB 704 PSTVVAENGYPAVELKDFSTDTKFL-----TFTNRTTHELY-----QMSDN 747  
QY 85 TDFNSLFTTTT---NGGRITVTVTKKDTASANEIATWAKSISANTTPVSTVTE----- 132  
DB 748 TDTNAVITSATDPNSGVLYDKIDGAA-----IKAGSNITVPAGKTAIEFTLSLPKS 800  
QY 133 -----SNNDGTEVINVSQYGYVYSSTVNNGAVI-MVTSVTPNAT----- 171  
DB 801 FPOQGFVEGLNFKGSDGSR-LNLPYMGFF---GDWNDGKIVDSLNGITYSPAGNFGTV 856  
QY 172 --IHEKNTTATWGDG-----CGKTVQDKTYSVGDVTKVITYYKNAVNVHGTGKVVYVI 223  
DB 857 PLLKNKNTQTYGGMVTDAGNKTVDDQAIAPSSD-----KNAL-YNDISMKY-YLL 907  
QY 224 KDTMPSASVVDLNEGSYEVTITDGSN-ITTLTQSEKATGKYNLLENNFTIIPWA 282  
DB 908 R-----NISNVQVILDGQGNKVTLLSSSTNRKKTYNAHSQYIYNAPWDG 956  
QY 283 T--NTPTGNTGANDDPFYKGINITVTVYTVGLKSGAKPGSADLPENTINATNPNTSN 340  
DB 957 TYDQDQGNIKTADDSVYR-----ISGVPEGGDKRQVDFVPFKLD----- 998  
QY 341 DPGQKTVTVRQGITIKIDGSTK-----ASLQGAFLVKNATGQFLNPDNNVWGTGA 396  
DB 999 ---SKAPTVRHVALSAKTENGKTQVYLTAAEKDDLSGLDATKSVKTEINEVTNL-----DA 1051

QY 397 NATEYTTGAGIITI-TGLKEGTYYLVEKKAPLYNLLNSOKVILGCGATDTTNSDNL 455  
ID YFJA ECOLI STANDARD; PRT; 1569 AA.  
DB 1052 TPTDAGTAGYKIEPLSD-----EQAALGNG--DNSAELYLTDNASATDQASV 1103  
QY 456 VNPTVENNKTELPSTGGIGTITFYIIGALVIGAGI 492  
DB 1104 QKP-----GSTSFDLI-----VNGGGI 1120

## RESULT 18

YFJA ECOLI  
AC P52143; P77017; P77019; PRT; 1569 AA.  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical outer membrane protein YPJA.  
GN YFJA OR B2647.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R. Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner K.J., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RA "The complete genome sequence of Escherichia coli K-12";  
RN Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=K12; PubMed=9205837;  
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,  
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horiuchi T.;  
RA "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT analysis of its sequence features";  
RL DNA Res. 4:91-113(1997).  
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).  
CC -!- SIMILARITY: STRONG, TO BORDETTELLA PERTACTIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

CC -----  
DR EMBL; U36840; AAA79815.1; ALT\_SEQ.  
DR EMBL; AB000350; AAC75695.1; -.  
DR EMBL; D90889; BAA16514.1; ALT\_INIT.  
DR EMBL; D90890; BAA16518.1; ALT\_INIT.  
DR PIR; A65044; A65044.  
DR EcoGene; EG13213; ypJA.  
DR InterPro; IPR006315; Autotransport.  
DR InterPro; IPR005546; Autotransporter.  
DR InterPro; IPR004899; Pertactin.  
DR InterPro; IPR003991; Pertactin C.  
DR Pfam; PF03797; Autotransporter; 1.  
DR Pfam; PF03212; Pertactin; 1.  
DR PRINTS; PR01484; PRCTACNFAMLY.  
DR TIGRFAMs; TIGR01414; Autotrans barl; 2.  
KW Hypothetical protein; Outer membrane; Complete proteome.

SQ SEQUENCE 1569 AA; 162774 MW; 0D006A9A32154596 CRC64;  
Query Match 6.4%; Score 164; DB 1; Length 1569;  
Best Local Similarity 25.3%; Pred. No. 0.13;  
Matches 137; Conservative 55; Mismatches 208; Indels 142; Gaps 30;  
QY 32 TGTITV---QDTQXG--ATYKAYKVFDAIDNANVSNKDGASYLI-----PQKEAE 80  
DB 159 TGT-TIESGNQDVYKGGISNGTTIKGASRVGGSGANGILIDGGQIVKVGQHDGTTIN 217  
QY 81 YKASTDF--NSLFTTTT-NGRTYV---TKKDTASANEIATWAKSISANTTTPVSTVTESN 134  
DB 218 KSGSQDVWQGLATNTTNGGQYVEQSTVETTTKNGGEQRVYESRALDTTIEGQTSL 277  
QY 135 NDGTEVINVSQY--GYVYVST-----VNGGAVIMTSTVTPNATIEKNTDATTWGDG 185  
DB 278 NKSSTAKNTHIVSGGTQIVDNTSTSDVIEVSGVLDVRGGTA-----TNVQHDS- 328  
QY 186 GKTVPQKTVSGDTVKYITTYKNAVNYHTEKVKYQYVIKDT-MPSASVVDLNE-GSYEVT 243  
DB 329 ---AILKTNNGTTVSGT-----NSEGAFSIHNVHADNVLLENGCHLDINAYGSANKT 378  
QY 244 IYDGSNITTLTQSGEKATKYNLLENNNTTITPWAATNT-PTGNTONGANDDFFYKG 302  
DB 379 IIKDKGTSVLTNKADAT-----RIDNGGYMDVAGNATNTIINGGTQINNN-----YG 427  
QY 303 INTITVTVYTG--LKSGAK-----PGSADLPENTNTIATINPNTS----- 339  
DB 428 IATGNTNNGTQNIKSGGKADTTIISGSRQVVEKDGTA-IGSNISAGSLIVYTGIAH 486  
QY 340 --NDDPGQKTVRDRGQITIKKIDGSTKAS---LOG--AIFVLKNATGQF----- 381  
DB 487 GVNQETGSALVANTGAGT--DIEGYNKLSHFTITGGEANYVVLNTGELTVVAKTSAKNT 544  
QY 382 -----LNFDNINNVETWEATNATYTTGADGIITG--LKEGT- 418  
DB 545 TIDTGKLIQVKEAKTDTSLNNGGVLEVDQGEAKHVEQQSGGALLIATTSGLTLEGTN 604  
QY 419 -----YYLVEKKAPLYNLLDN--SOKVILGCGATDTTNSDNLVN-----PTVENNKG 465  
DB 605 SYGDAPYIRNSEAK--NVLENAGSLTVVTGSRVAVDTIINANGKMDVYKDVGVTLNSAG 662  
QY 466 TE 467  
DB 663 TQ 664

## RESULT 19

BIGA\_SALTY STANDARD; PRT; 1953 AA.  
ID BIGA\_SALTY STANDARD; PRT; 1953 AA.  
AC P25927; P25928; Q9XCO3;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Putative surface-exposed virulence protein bigA precursor.  
GN BIGA OR STW3478.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 14028;  
RA Stojiljkovic I., Valentine P., Heffron F.;  
RT "Salmonella typhimurium rhs homolog.";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,







RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi A., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vanier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*,"  
 RL Nature 390:249-256(1997).  
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,  
 CC MOTILITY, SECRETION OR DIFFERENTIATION.  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED  
 CC INTO THE MEDIUM.  
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE  
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE  
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED  
 CC MOTIF REPEATED 31 TIMES.  
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME  
 CC SIMILARITY TO THE REPEAT IN E. COLI RHS GROUP OF PROTEINS (RHS-A-D).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L05634; AAA22893.1; -;  
 DR EMBL; D31856; BAA06656.1; -;  
 DR EMBL; D29985; BAA06260.1; -;  
 DR EMBL; D83026; BAA11683.1; -;  
 DR EMBL; Z99124; CAB15959.1; -;  
 DR PIR; S32920; S32920.  
 DR Subtilast; BG10797; wapa.  
 DR InterPro; IPR003305; CBM\_CenC.  
 DR InterPro; IPR006530; YD.  
 DR Pfam; PF02018; CBM 4.9; 1.  
 DR Pfam; PF05593; RHS repeat; 14.  
 DR TIGRfam; TIGR01643; YD repeat 2x; 17.  
 KW Cell wall; Repeat; Signal; Complete proteome.  
 FT SIGNAL 1 28  
 FT CHAIN 29 2334  
 FT DOMAIN 504 869  
 FT REPEAT 504 605  
 FT REPEAT 636 736  
 FT REPEAT 769 869  
 FT DOMAIN 1021 2139  
 FT REPEAT 1021 1040  
 FT REPEAT 1042 1061  
 FT REPEAT 1063 1082  
 FT REPEAT 1083 1102  
 FT REPEAT 1109 1128  
 FT REPEAT 1129 1148  
 FT REPEAT 1150 1169  
 FT REPEAT 1174 1193  
 FT REPEAT 1199 1218  
 FT REPEAT 1219 1238  
 FT REPEAT 1246 1265  
 FT REPEAT 1267 1286  
 FT REPEAT 1287 1306  
 FT REPEAT 1307 1326  
 FT REPEAT 1327 1346  
 FT REPEAT 1347 1366  
 FT REPEAT 1367 1386  
 FT REPEAT 1387 1406  
 FT REPEAT 1407 1426  
 FT REPEAT 1427 1446  
 FT REPEAT 1447 1466  
 FT REPEAT 1467 1486  
 FT REPEAT 1487 1506  
 FT REPEAT 1507 1526  
 FT REPEAT 1527 1546  
 FT REPEAT 1547 1566  
 FT REPEAT 1567 1586  
 FT REPEAT 1587 1606  
 FT REPEAT 1607 1626  
 FT REPEAT 1627 1646  
 FT REPEAT 1647 1666  
 FT REPEAT 1667 1686  
 FT REPEAT 1687 1706  
 FT REPEAT 1707 1726  
 FT REPEAT 1727 1746  
 FT REPEAT 1747 1766  
 FT REPEAT 1767 1786  
 FT REPEAT 1787 1806  
 FT REPEAT 1807 1826  
 FT REPEAT 1827 1846  
 FT REPEAT 1847 1866  
 FT REPEAT 1867 1886  
 FT REPEAT 1887 1906  
 FT REPEAT 1907 1926  
 FT REPEAT 1927 1946  
 FT REPEAT 1947 1966  
 FT REPEAT 1967 1986  
 FT REPEAT 1987 2006  
 FT REPEAT 2007 2026  
 FT REPEAT 2027 2046  
 FT REPEAT 2047 2066  
 FT REPEAT 2067 2086  
 FT REPEAT 2087 2106  
 FT REPEAT 2107 2126  
 FT REPEAT 2127 2146  
 FT REPEAT 2147 2166  
 FT REPEAT 2167 2186  
 FT REPEAT 2187 2206  
 FT REPEAT 2207 2226  
 FT REPEAT 2227 2246  
 FT REPEAT 2247 2266  
 FT REPEAT 2267 2286  
 FT REPEAT 2287 2306  
 FT REPEAT 2307 2326  
 FT REPEAT 2327 2346  
 FT REPEAT 2347 2366  
 FT REPEAT 2367 2386  
 FT REPEAT 2387 2406  
 FT REPEAT 2407 2426  
 FT REPEAT 2427 2446  
 FT REPEAT 2447 2466  
 FT REPEAT 2467 2486  
 FT REPEAT 2487 2506  
 FT REPEAT 2507 2526  
 FT REPEAT 2527 2546  
 FT REPEAT 2547 2566  
 FT REPEAT 2567 2586  
 FT REPEAT 2587 2606  
 FT REPEAT 2607 2626  
 FT REPEAT 2627 2646  
 FT REPEAT 2647 2666  
 FT REPEAT 2667 2686  
 FT REPEAT 2687 2706  
 FT REPEAT 2707 2726  
 FT REPEAT 2727 2746  
 FT REPEAT 2747 2766  
 FT REPEAT 2767 2786  
 FT REPEAT 2787 2806  
 FT REPEAT 2807 2826  
 FT REPEAT 2827 2846  
 FT REPEAT 2847 2866  
 FT REPEAT 2867 2886  
 FT REPEAT 2887 2906  
 FT REPEAT 2907 2926  
 FT REPEAT 2927 2946  
 FT REPEAT 2947 2966  
 FT REPEAT 2967 2986  
 FT REPEAT 2987 3006  
 FT REPEAT 3007 3026  
 FT REPEAT 3027 3046  
 FT REPEAT 3047 3066  
 FT REPEAT 3067 3086  
 FT REPEAT 3087 3106  
 FT REPEAT 3107 3126  
 FT REPEAT 3127 3146  
 FT REPEAT 3147 3166  
 FT REPEAT 3167 3186  
 FT REPEAT 3187 3206  
 FT REPEAT 3207 3226  
 FT REPEAT 3227 3246  
 FT REPEAT 3247 3266  
 FT REPEAT 3267 3286  
 FT REPEAT 3287 3306  
 FT REPEAT 3307 3326  
 FT REPEAT 3327 3346  
 FT REPEAT 3347 3366  
 FT REPEAT 3367 3386  
 FT REPEAT 3387 3406  
 FT REPEAT 3407 3426  
 FT REPEAT 3427 3446  
 FT REPEAT 3447 3466  
 FT REPEAT 3467 3486  
 FT REPEAT 3487 3506  
 FT REPEAT 3507 3526  
 FT REPEAT 3527 3546  
 FT REPEAT 3547 3566  
 FT REPEAT 3567 3586  
 FT REPEAT 3587 3606  
 FT REPEAT 3607 3626  
 FT REPEAT 3627 3646  
 FT REPEAT 3647 3666  
 FT REPEAT 3667 3686  
 FT REPEAT 3687 3706  
 FT REPEAT 3707 3726  
 FT REPEAT 3727 3746  
 FT REPEAT 3747 3766  
 FT REPEAT 3767 3786  
 FT REPEAT 3787 3806  
 FT REPEAT 3807 3826  
 FT REPEAT 3827 3846  
 FT REPEAT 3847 3866  
 FT REPEAT 3867 3886  
 FT REPEAT 3887 3906  
 FT REPEAT 3907 3926  
 FT REPEAT 3927 3946  
 FT REPEAT 3947 3966  
 FT REPEAT 3967 3986  
 FT REPEAT 3987 4006  
 FT REPEAT 4007 4026  
 FT REPEAT 4027 4046  
 FT REPEAT 4047 4066  
 FT REPEAT 4067 4086  
 FT REPEAT 4087 4106  
 FT REPEAT 4107 4126  
 FT REPEAT 4127 4146  
 FT REPEAT 4147 4166  
 FT REPEAT 4167 4186  
 FT REPEAT 4187 4206  
 FT REPEAT 4207 4226  
 FT REPEAT 4227 4246  
 FT REPEAT 4247 4266  
 FT REPEAT 4267 4286  
 FT REPEAT 4287 4306  
 FT REPEAT 4307 4326  
 FT REPEAT 4327 4346  
 FT REPEAT 4347 4366  
 FT REPEAT 4367 4386  
 FT REPEAT 4387 4406  
 FT REPEAT 4407 4426  
 FT REPEAT 4427 4446  
 FT REPEAT 4447 4466  
 FT REPEAT 4467 4486  
 FT REPEAT 4487 4506  
 FT REPEAT 4507 4526  
 FT REPEAT 4527 4546  
 FT REPEAT 4547 4566  
 FT REPEAT 4567 4586  
 FT REPEAT 4587 4606  
 FT REPEAT 4607 4626  
 FT REPEAT 4627 4646  
 FT REPEAT 4647 4666  
 FT REPEAT 4667 4686  
 FT REPEAT 4687 4706  
 FT REPEAT 4707 4726  
 FT REPEAT 4727 4746  
 FT REPEAT 4747 4766  
 FT REPEAT 4767 4786  
 FT REPEAT 4787 4806  
 FT REPEAT 4807 4826  
 FT REPEAT 4827 4846  
 FT REPEAT 4847 4866  
 FT REPEAT 4867 4886  
 FT REPEAT 4887 4906  
 FT REPEAT 4907 4926  
 FT REPEAT 4927 4946  
 FT REPEAT 4947 4966  
 FT REPEAT 4967 4986  
 FT REPEAT 4987 5006  
 FT REPEAT 5007 5026  
 FT REPEAT 5027 5046  
 FT REPEAT 5047 5066  
 FT REPEAT 5067 5086  
 FT REPEAT 5087 5106  
 FT REPEAT 5107 5126  
 FT REPEAT 5127 5146  
 FT REPEAT 5147 5166  
 FT REPEAT 5167 5186  
 FT REPEAT 5187 5206  
 FT REPEAT 5207 5226  
 FT REPEAT 5227 5246  
 FT REPEAT 5247 5266  
 FT REPEAT 5267 5286  
 FT REPEAT 5287 5306  
 FT REPEAT 5307 5326  
 FT REPEAT 5327 5346  
 FT REPEAT 5347 5366  
 FT REPEAT 5367 5386  
 FT REPEAT 5387 5406  
 FT REPEAT 5407 5426  
 FT REPEAT 5427 5446  
 FT REPEAT 5447 5466  
 FT REPEAT 5467 5486  
 FT REPEAT 5487 5506  
 FT REPEAT 5507 5526  
 FT REPEAT 5527 5546  
 FT REPEAT 5547 5566  
 FT REPEAT 5567 5586  
 FT REPEAT 5587 5606  
 FT REPEAT 5607 5626  
 FT REPEAT 5627 5646  
 FT REPEAT 5647 5666  
 FT REPEAT 5667 5686  
 FT REPEAT 5687 5706  
 FT REPEAT 5707 5726  
 FT REPEAT 5727 5746  
 FT REPEAT 5747 5766  
 FT REPEAT 5767 5786  
 FT REPEAT 5787 5806  
 FT REPEAT 5807 5826  
 FT REPEAT 5827 5846  
 FT REPEAT 5847 5866  
 FT REPEAT 5867 5886  
 FT REPEAT 5887 5906  
 FT REPEAT 5907 5926  
 FT REPEAT 5927 5946  
 FT REPEAT 5947 5966  
 FT REPEAT 5967 5986  
 FT REPEAT 5987 6006  
 FT REPEAT 6007 6026  
 FT REPEAT 6027 6046  
 FT REPEAT 6047 6066  
 FT REPEAT 6067 6086  
 FT REPEAT 6087 6106  
 FT REPEAT 6107 6126  
 FT REPEAT 6127 6146  
 FT REPEAT 6147 6166  
 FT REPEAT 6167 6186  
 FT REPEAT 6187 6206  
 FT REPEAT 6207 6226  
 FT REPEAT 6227 6246  
 FT REPEAT 6247 6266  
 FT REPEAT 6267 6286  
 FT REPEAT 6287 6306  
 FT REPEAT 6307 6326  
 FT REPEAT 6327 6346  
 FT REPEAT 6347 6366  
 FT REPEAT 6367 6386  
 FT REPEAT 6387 6406  
 FT REPEAT 6407 6426  
 FT REPEAT 6427 6446  
 FT REPEAT 6447 6466  
 FT REPEAT 6467 6486  
 FT REPEAT 6487 6506  
 FT REPEAT 6507 6526  
 FT REPEAT 6527 6546  
 FT REPEAT 6547 6566  
 FT REPEAT 6567 6586  
 FT REPEAT 6587 6606  
 FT REPEAT 6607 6626  
 FT REPEAT 6627 6646  
 FT REPEAT 6647 6666  
 FT REPEAT 6667 6686  
 FT REPEAT 6687 6706  
 FT REPEAT 6707 6726  
 FT REPEAT 6727 6746  
 FT REPEAT 6747 6766  
 FT REPEAT 6767 6786  
 FT REPEAT 6787 6806  
 FT REPEAT 6807 6826  
 FT REPEAT 6827 6846  
 FT REPEAT 6847 6866  
 FT REPEAT 6867 6886  
 FT REPEAT 6887 6906  
 FT REPEAT 6907 6926  
 FT REPEAT 6927 6946  
 FT REPEAT 6947 6966  
 FT REPEAT 6967 6986  
 FT REPEAT 6987 7006  
 FT REPEAT 7007 7026  
 FT REPEAT 7027 7046  
 FT REPEAT 7047 7066  
 FT REPEAT 7067 7086  
 FT REPEAT 7087 7106  
 FT REPEAT 7107 7126  
 FT REPEAT 7127 7146  
 FT REPEAT 7147 7166  
 FT REPEAT 7167 7186  
 FT REPEAT 7187 7206  
 FT REPEAT 7207 7226  
 FT REPEAT 7227 7246  
 FT REPEAT 7247 7266  
 FT REPEAT 7267 7286  
 FT REPEAT 7287 7306  
 FT REPEAT 7307 7326  
 FT REPEAT 7327 7346  
 FT REPEAT 7347 7366  
 FT REPEAT 7367 7386  
 FT REPEAT 7387 7406  
 FT REPEAT 7407 7426  
 FT REPEAT 7427 7446  
 FT REPEAT 7447 7466  
 FT REPEAT 7467 7486  
 FT REPEAT 7487 7506  
 FT REPEAT 7507 7526  
 FT REPEAT 7527 7546  
 FT REPEAT 7547 7566  
 FT REPEAT 7567 7586  
 FT REPEAT 7587 7606  
 FT REPEAT 7607 7626  
 FT REPEAT 7627 7646  
 FT REPEAT 7647 7666  
 FT REPEAT 7667 7686  
 FT REPEAT 7687 7706  
 FT REPEAT 7707 7726  
 FT REPEAT 7727 7746  
 FT REPEAT 7747 7766  
 FT REPEAT 7767 7786  
 FT REPEAT 7787 7806  
 FT REPEAT 7807 7826  
 FT REPEAT 7827 7846  
 FT REPEAT 7847 7866  
 FT REPEAT 7867 7886  
 FT REPEAT 7887 7906  
 FT REPEAT 7907 7926  
 FT REPEAT 7927 7946  
 FT REPEAT 7947 7966  
 FT REPEAT 7967 7986  
 FT REPEAT 7987 8006  
 FT REPEAT 8007 8026  
 FT REPEAT 8027 8046  
 FT REPEAT 8047 8066  
 FT REPEAT 8067 8086  
 FT REPEAT 8087 8106  
 FT REPEAT 8107 8126  
 FT REPEAT 8127 8146  
 FT REPEAT 8147 8166  
 FT REPEAT 8167 8186  
 FT REPEAT 8187 8206  
 FT REPEAT 8207 8226  
 FT REPEAT 8227 8246  
 FT REPEAT 8247 8266  
 FT REPEAT 8267 8286  
 FT REPEAT 8287 8306  
 FT REPEAT 8307 8326  
 FT REPEAT 8327 8346  
 FT REPEAT 8347 8366  
 FT REPEAT 8367 8386  
 FT REPEAT 8387 8406  
 FT REPEAT 8407 8426  
 FT REPEAT 8427 8446  
 FT REPEAT 8447 8466  
 FT REPEAT 8467 8486  
 FT REPEAT 8487 8506  
 FT REPEAT 8507 8526  
 FT REPEAT 8527 8546  
 FT REPEAT 8547 8566  
 FT REPEAT 8567 8586  
 FT REPEAT 8587 8606  
 FT REPEAT 8607 8626  
 FT REPEAT 8627 8646  
 FT REPEAT 8647 8666  
 FT REPEAT 8667 8686  
 FT REPEAT 8687 8706  
 FT REPEAT 8707 8726  
 FT REPEAT 8727 8746  
 FT REPEAT 8747 8766  
 FT REPEAT 8767 8786  
 FT REPEAT 8787 8806  
 FT REPEAT 8807 8826  
 FT REPEAT 8827 8846  
 FT REPEAT 8847 8866  
 FT REPEAT 8867 8886  
 FT REPEAT 8887 8906  
 FT REPEAT 8907 8926  
 FT REPEAT 8927 8946  
 FT REPEAT 8947 8966  
 FT REPEAT 8967 8986  
 FT REPEAT 8987 9006  
 FT REPEAT 9007 9026  
 FT REPEAT 9027 9046  
 FT REPEAT 9047 9066  
 FT REPEAT 9067 9086  
 FT REPEAT 9087 9106  
 FT REPEAT 9107 9126  
 FT REPEAT 9127 9146  
 FT REPEAT 9147 9166  
 FT REPEAT 9167 9186  
 FT REPEAT 9187 9206  
 FT REPEAT 9207 9226  
 FT REPEAT 9227 9246  
 FT REPEAT 9247 9266  
 FT REPEAT 9267 9286  
 FT REPEAT 9287 9306  
 FT REPEAT 9307 9326  
 FT REPEAT 9327 9346  
 FT REPEAT 9347 9366  
 FT REPEAT 9367 9386  
 FT REPEAT 9387 9406  
 FT REPEAT 9407 9426  
 FT REPEAT 9427 9446  
 FT REPEAT 9447 9466  
 FT REPEAT 9467 9486  
 FT REPEAT 9487 9506  
 FT REPEAT 9507 9526  
 FT REPEAT 9527 9546  
 FT REPEAT 9547 9566  
 FT REPEAT 9567 9586  
 FT REPEAT 9587 9606  
 FT REPEAT 9607 9626  
 FT REPEAT 9627 9646  
 FT REPEAT 9647 9666  
 FT REPEAT 9667 9686  
 FT REPEAT 9687 9706  
 FT REPEAT 9707 9726  
 FT REPEAT 9727 9746  
 FT REPEAT 9747 9766  
 FT REPEAT 9767 9786  
 FT REPEAT 9787 9806  
 FT REPEAT 9807 9826  
 FT REPEAT 9827 9846  
 FT REPEAT 9847 9866  
 FT REPEAT 9867 9886  
 FT REPEAT 9887 9906  
 FT REPEAT 9907 9926  
 FT REPEAT 9927 9946  
 FT REPEAT 9947 9966  
 FT REPEAT 9967 9986  
 FT REPEAT 9987 10006  
 FT REPEAT 10007 10026  
 FT REPEAT 10027 10046  
 FT REPEAT 10047 10066  
 FT REPEAT 10067 10086  
 FT REPEAT 10087 10106  
 FT REPEAT 10107 10126  
 FT REPEAT 10127 10146  
 FT REPEAT 10147 10166  
 FT REPEAT 10167 10186  
 FT REPEAT 10187 10206  
 FT REPEAT 10207 10226  
 FT REPEAT 10227 10246  
 FT REPEAT 10247 10266  
 FT REPEAT 10267 10286  
 FT REPEAT 10287 10306  
 FT REPEAT 10307 10326  
 FT REPEAT 10327 10346  
 FT REPEAT 10347 10366  
 FT REPEAT 10367 10386  
 FT REPEAT 10387 10406  
 FT REPEAT 10407 10426  
 FT REPEAT 10427 10446  
 FT REPEAT 10447 10466  
 FT REPEAT 10467 10486  
 FT REPEAT 10487 10506  
 FT REPEAT 10507 10526  
 FT REPEAT 10527 10546  
 FT REPEAT 10547 10566  
 FT REPEAT 10567 10586  
 FT REPEAT 10587 10606  
 FT REPEAT 10607 10626  
 FT REPEAT 10627 10646  
 FT REPEAT 10647 10666  
 FT REPEAT 10667 10686  
 FT REPEAT 10687 10706  
 FT REPEAT 10707 10726  
 FT REPEAT 10727 10746  
 FT REPEAT 10747 10766  
 FT REPEAT 10767 10786  
 FT REPEAT 10787 10806  
 FT REPEAT 10807 10826  
 FT REPEAT 10827 10846  
 FT REPEAT 10847 10866  
 FT REPEAT 10867 10886  
 FT REPEAT 10887 10906  
 FT REPEAT 10907 10926  
 FT REPEAT 10927 10946  
 FT REPEAT 10947 10966  
 FT REPEAT 10967 10986  
 FT REPEAT 10987 11006  
 FT REPEAT 11007 11026  
 FT REPEAT 11027 11046  
 FT REPEAT 11047 11066  
 FT REPEAT 11067 11086  
 FT REPEAT 11087 11106  
 FT REPEAT 11107 11126  
 FT REPEAT 11127 11146  
 FT REPEAT 11147 11166  
 FT REPEAT 11167 11186  
 FT REPEAT 11187 11206  
 FT REPEAT 11207 11226  
 FT REPEAT 11227 11246  
 FT REPEAT 11247 11266  
 FT REPEAT 11267 11286  
 FT REPEAT 11287 11306  
 FT REPEAT 11307 11326  
 FT REPEAT 11327 11346  
 FT REPEAT 11347 11366  
 FT REPEAT 11367 11386  
 FT REPEAT 11387 11406  
 FT REPEAT 11407 11426  
 FT REPEAT 11427 11446  
 FT REPEAT 11447 11466  
 FT REPEAT 11467 11486  
 FT REPEAT 11487 11506  
 FT REPEAT 11507 11526  
 FT REPEAT 11527 11546  
 FT REPEAT 11547 11566  
 FT REPEAT 11567 11586  
 FT REPEAT 11587 11606  
 FT REPEAT 11607 11626  
 FT REPEAT 11627 11646  
 FT REPEAT 11647 11666  
 FT REPEAT 11667 11686  
 FT REPEAT 11687 11706  
 FT REPEAT 11707 11726  
 FT REPEAT 11727 11746  
 FT REPEAT 11747 11766  
 FT REPEAT 11767 11786  
 FT REPEAT 11787 11806  
 FT REPEAT 11807 11826  
 FT REPEAT 11827 11846  
 FT REPEAT 11847 11866  
 FT REPEAT 11867 11886  
 FT REPEAT 11887 11906  
 FT REPEAT 11907 11926  
 FT REPEAT 11927 11946  
 FT REPEAT 11947 11966  
 FT REPEAT 11967 11986  
 FT REPEAT 11987 12006  
 FT REPEAT 12007 12026  
 FT REPEAT 12027 12046  
 FT REPEAT 12047 12066  
 FT REPEAT 12067 12086  
 FT REPEAT 12087 12106  
 FT REPEAT 12107 12126  
 FT REPEAT 12127 12146  
 FT REPEAT 12147 12166  
 FT REPEAT 12167 12186  
 FT REPEAT 12187 12206  
 FT REPEAT 12207 12226  
 FT REPEAT 12227 12246  
 FT REPEAT 12247 12266  
 FT REPEAT 12267 12286  
 FT REPEAT 12287 12306  
 FT REPEAT 12307 12326  
 FT REPEAT 12327 12346  
 FT REPEAT 12347 12366  
 FT REPEAT 12367 12386  
 FT REPEAT 12387 12406  
 FT REPEAT 12407 12426  
 FT REPEAT 12427 12446  
 FT REPEAT 12447 12466  
 FT REPEAT 12467 12486  
 FT REPEAT 12487 12506  
 FT REPEAT 12507 12526  
 FT REPEAT 12527 12546  
 FT REPEAT 12547 12566  
 FT REPEAT 12567 12586  
 FT REPEAT 12587 12606  
 FT REPEAT 12607 12626  
 FT REPEAT 12627 12646  
 FT REPEAT 12647 12666  
 FT REPEAT 12667 12686  
 FT REPEAT 12687 12706  
 FT REPEAT 12707 12726  
 FT REPEAT 12727 12746  
 FT REPEAT 12747 12766  
 FT REPEAT 12767 12786  
 FT REPEAT 12787 12806  
 FT REPEAT 12807 12826  
 FT REPEAT 12827 12846  
 FT REPEAT 12847 12866  
 FT REPEAT 12867 12886  
 FT REPEAT 12887 12906  
 FT REPEAT 12907 12926  
 FT REPEAT 12927 12946  
 FT REPEAT 12947 12966  
 FT REPEAT 12967 12986  
 FT REPEAT 12987 13006  
 FT REPEAT 13007 13026  
 FT REPEAT 13027 13046  
 FT REPEAT 13047 13066  
 FT REPEAT 13067 13086  
 FT REPEAT 13087 13106  
 FT REPEAT 13107 13126  
 FT REPEAT 13127 13146  
 FT REPEAT 13147 13166  
 FT REPEAT 13167 13186  
 FT REPEAT 13187 13206  
 FT REPEAT 13207 13226  
 FT REPEAT 13227 13246  
 FT REPEAT 13247 13266  
 FT REPEAT 13267 13286  
 FT REPEAT 13287 13306  
 FT REPEAT 13307 13326  
 FT REPEAT 13327 13346  
 FT REPEAT 13347 13366  
 FT REPEAT 13367 13386  
 FT REPEAT 13387 13406  
 FT REPEAT 13407 13426  
 FT REPEAT 13427 13446  
 FT REPEAT 13447 13466  
 FT REPEAT 13467 13486  
 FT REPEAT 13487 13506  
 FT REPEAT 13507 13526  
 FT REPEAT 13527 13546  
 FT REPEAT 13547 13566  
 FT REPEAT 13567 13586  
 FT REPEAT 13587 13606  
 FT REPEAT 13607 13626  
 FT REPEAT 1362

RL J. Bacteriol. 168:365-373(1986).  
RN [2]  
RP SEQUENCE OF 1-85 FROM N.A.  
RC STRAIN=47;  
RX MEDLINE=88115203; PubMed=2828336;  
RA Tauboi A., Uchihi R., Adachi T., Sasaki T., Hayakawa S., Yamagata H.,  
RA Teukagoshi N., Uda S.;  
RT "Characterization of the genes for the hexagonally arranged surface  
RT layer proteins in protein-producing Bacillus brevis 47: complete  
RT nucleotide sequence of the middle wall protein gene."  
RL J. Bacteriol. 170:935-945(1988).  
CC -I- FUNCTION: THE OUTER WALL PROTEIN BINDS TO THE MIDDLE CELL WALL  
CC PROTEIN.  
CC -I- SUBUNIT: THE OUTER CELL WALL LAYER IS COMPOSED OF SUBUNITS OF  
CC THE OUTER CELL WALL PROTEIN. THESE PROTEINS FORM A HEXAGONAL  
CC ARRAY WITH A LATTICE CONSTANT OF 14.5 NM IN THE OUTER CELL  
CC WALL LAYERS.  
CC -I- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-  
CC layer with hexagonal symmetry.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; M14238; AAA22373.1; -  
DR EMBL; M19115; AAA22761.1; -  
DR PIR; B25039; B25039.  
KW Cell wall; S-layer; Signal.  
FT SIGNAL 1 24  
FT CHAIN 25 1004 OUTER CELL WALL PROTEIN  
FT SEQUENCE 1004 AA; 106137 MW; 3534F65BDD0ED9E CRC64;  
SQ  
  
Query Match 6.1%; Score 158; DB 1; Length 1004;  
Best Local Similarity 20.8%; Pred. No. 0.16;  
Matches 130; Conservative 75; Mismatches 202; Indels 218; Gaps 27;  
  
Qy 16 FGMVSPVPIFAAETGT---ITVQTKGATYKAYKVFDAIDNANVSDSNKDGASYL 72  
Db 412 FAADTPPTTKVBAKNTNTHVTFSETVAGADKANFTLKGVGN-----V 458  
Qy 73 IPQCKEAEYKASTDFNSLFTT-TNNGRTVYTKK--DTASANEIATWAKSIS-ANTTPVS 128  
Db 459 IPLTKAEVDAAKNIYKVVTPEPLNGGSYLTVKGIEDASKNKLVEYTAIVADIVPPN 518  
Qy 129 -----TVTESNN 135  
Db 519 VKLDLPATPGTDAQIISPTKVKIAFTPEPMKASIKNNYMFNGFNLDKSVTLTATDSNT 578  
Qy 136 ----DGEVINVSQYGYYSSTVNGAVIMVTSVTPNATHKEKNTDWTGDSGGKTVQD 191  
Db 579 AVVVDFTNVVGFNGF-----KNGDAISVGRVLDTA-----GNPKTEMQ 616  
Qy 192 KTVSGDVTVKTIITIKNAVNYHGTKEY-QYVIKDTMPSASVVDLNEGSYE----- 241  
Db 617 TKVNLPSVSAFLPDKAEVTKNIVKLYFKELIINAKADDPVNDGEG-YKAVNSISNDV 675  
Qy 242 -----VTIDSGNITTLTQSGKATGKYNLLENNFTIIPWATNTP----- 286  
Db 676 VENKSVITLTGNDLPPTTAAGVKVKTGVE---VDKKNQYGVAV---ALTDVPADDKIGPNW 730  
Qy 287 ----TGNTONGANDDPFYKGINITITVTVTGVLKSGAKPGSADLPNTNIAINPTNSDD 342  
Db 731 LKAETVDTNNGKIDQF-----KLTFSEALYVASVQDSFRIEYTIAGV-----ET 777  
Qy 343 PQQKTVTRDGGQITIKKIDGSTKASIQGAIFVLK-NATQGF-----LNFND-----TN 388  
Db 778 KGEVTVIKVTELDIDSDATPVAIGSVGVDLKRNASGPFEPQKAIKDVSAAPKEAPVWT 837  
Qy 389 NVEWTEANATEYTTGADGIITITGLKE-----GTYILV----- 422

Db 838 GVEAGKTYNTAVTPDSADKDIKTVLKKDQKGLAGYALKTPISNGSYVELVVDNAGNNT 897  
Qy 423 -----EKKAPLGYNLLDNSQKVLGDG-----ATDTN---SDNLLVNPT 459  
Db 898 TVKFKVDIPADSKKAPKTEIKTVD--KVAVADAPKWEAPKATATDDVDGDISDKIATVYS 955  
Qy 460 VEN--NKGTELPST---GGIGTTI 478  
Db 956 SEDAGSKVTDLASAQTHLGTAGNTV 980  
  
RESULT 23  
AG43 ECOLI  
ID AG43 ECOLI STANDARD; PRT: 1039 AA.  
AC P39180; P75614; P76360; P97241; Q46771;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Antigen 43 precursor (AG43) (Fluffing protein).  
GN FLU OR B2000.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12."  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97251358; PubMed=9097040;  
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,  
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,  
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,  
RA Sivasubraman S., Tagami H., Takeda J., Takemoto K., Wada C.,  
RA Yamanoto Y., Horuchi T.;  
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome  
RT corresponding to the 40.1-50.0 min region on the linkage map."  
RL DNA Res. 3:379-392(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ML 308-225;  
RA Henderson I.R., Owen P.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP PRELIMINARY SEQUENCE OF 53-78.  
RC STRAIN=ML 308-225;  
RX MEDLINE=89291704; PubMed=2661530;  
RA Caffrey P., Owen P.;  
RT "Purification and N-terminal sequence of the alpha subunit of antigen  
RT 43, a unique protein complex associated with the outer membrane of  
RT Escherichia coli."  
RL J. Bacteriol. 171:3634-3640(1989).  
RN [5]  
RP SEQUENCE OF 53-63.  
RC STRAIN=K12 / EMG2;  
RX MEDLINE=97443975; PubMed=9298646;  
RA Link A.J., Robison K., Church G.M.;  
RT "Comparing the predicted and observed properties of proteins encoded  
RT in the genome of Escherichia coli K-12."  
RL Electrophoresis 18:1259-1313(1997).  
RN [6]  
RP GENE NAME.  
RX MEDLINE=97257509; PubMed=9103983;





Db 549 TGAATGTTTAAKPTGANGTITTTTAKPAGANGTITTTTAKPAGA 593

RESULT 26  
 OMPB RICJA STANDARD; PRT; 1656 AA.  
 AC 00653;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas) (rOMPb)  
 DE (rOMP B) (Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).  
 OS OMPB.  
 GN Rickettsia japonica.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=35790;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YH;  
 RA Uchiyama T.;  
 RT "Sequencing of the gene encoding the protein rOMP B of Rickettsia japonica."  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: The 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).  
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.  
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB003681; BAA20138.1; --  
 CC InterPro; IPR006315; Autotransport.  
 CC InterPro; IPR005546; Autotransporter.  
 CC Pfam; PF03797; Autotransporter; 1.  
 CC TIGRFAMS; TIGR01414; autotrans\_bar1; 2.  
 CC Antigen; S-layer; Cell wall.  
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.  
 FT DOMAIN 528 533 POLY-GLY.  
 FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;  
 SQ  
 Query Match 6.1%; Score 157.5; DB 1; Length 1656;  
 Best Local Similarity 22.5%; Pred. No. 0.32;  
 Matches 136; Conservative 73; Mismatches 217; Indels 179; Gaps 35;  
 Qy 12 ASLAFGMVSPVT---PIAFA-----AETGTVT-----QDTQK-ATYKAYKVPDAE 55  
 Db 182 STLVFDLA-NPTTQKAPLILADNALIVNGANGLVNTGFIQVSKSPATVKAINIGDQ 240  
 Qy 56 --IDNAVSDSN-----KDGASYLIPQCKEAEYKASTDFNSLFTTTINGG 98  
 Db 241 GFPMFNATNANALINQAGGTINFGTGTGRLLVSKNG---AATDFN--VTGSLGNN 295  
 Qy 99 RTVVTKKDTASANEIATWAKSISANTTPVSTVTSNNDGTE-----VINV----- 143  
 Db 296 LKGIIELTVAIN-----GQLIANAGPANAVICTNGAGRAAGFVSDVNGKAAIDGQ 349  
 Qy 144 -----SQGYGYYSSTVNGNAVIMVTSVTPNATIHEKNTDATWGDGGKTV 189

Db 350 VYAKDMVLOSANANGQVNFRIHVDVGIDGTTAFKTAASIVAITQNSFGTT--DFGNLAA 407

Qy 190 DOKTYSVGDIVKYTIITYKNAVNYHG--TEKYVQYVYKIDTSPASVVDLNEG-SYEVITIDG 247

Db 408 Q---VTVPDWTMTLGNFTGDANNPNTAGVITPAANGTCLASAS-ADANVAVTNNTAIEA 463

Qy 248 SGNITTLTGSEKATGCKYNNLEENNNFTIIPWAAATPTPGTONTONGANDDFYKGINITIT 307

Db 464 SG--GVGVQLSGTHTAELRLGNAGSVFKLA-----DGTVINGK-----VNO-T 503

Qy 308 VYTVGVLSGSA--KPGSADLPENTNIATINPNTSDDPG---QKVTVRDQGITIKKIDGS 362

Db 504 VLVGVVLAAGAITLDGS-----ATITGIDNGGGGAALQSITLAN-----DAT 546

Qy 363 TKASLQGAIFVLKNATGQFLNF-----NDTNVNEWGEA 396

Db 547 KTLTLGGANII--SANGGTINFGANGGTTKLTSTONNIVVDCDLAIATDQTVVDASSLT 604

Qy 397 NATEYT-TGADGIITITLKEGTYYLVEKAPL-GYNLLDNSQKVLIDGSA----- 445

Db 605 NAGTLTISGTIGIIGANNNTTLGQFNIGSKTTLNGNVAIN--ELVTGNNGSVQFAHTY 662

Qy 446 --TDTTNS---DNLLVNPVTENN---KGTELPSTGG-IGTTFIYIIGA---ILVIGAG 491

Db 663 LITRTNAGQKIIIFNVVNNNTTLAAGTNLGSAAANPLAEINFGSKGARADTVLVNVEG 722

Qy 492 IVLVA 496

Db 723 VNLXA 727

RESULT 27  
 SLAP LACHE STANDARD; PRT; 439 AA.  
 ID SLAP LACHE STANDARD; PRT; 439 AA.  
 AC P38059;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE S-layer protein precursor (Surface layer protein).  
 GN SLPH.  
 OS Lactobacillus helveticus.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=1587;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CNRZ 892, and CNRZ 1269;  
 RA Callegari M.L., Cocconcelli P.S., Morelli L.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN 12)  
 RP SEQUENCE OF 31-53.  
 RC STRAIN=ATCC 12046;  
 RA Lortal S., van Heijenoort J., Gruber K., Sleytr U.B.;  
 RT "S-layer of Lactobacillus helveticus ATCC 12046: isolation, chemical characterization and re-formation after extraction with lithium chloride."  
 RT J. Gen. Microbiol. 138:611-618(1992).  
 CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly of proteins which coat the surface of bacteria.  
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X91199; CA62606.1; --  
 CC EMBL; X92752; CA63409.1; --

Db 549 TGAATGTTTAAKPTGANGTITTTTAKPAGANGTITTTTAKPAGA 593

RESULT 26  
 OMPB RICJA STANDARD; PRT; 1656 AA.

AC O06653; 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)  
 DE (Surface protein antigen) (Cell surface antigen 5) (Scas5) (rOMPb)  
 DE (rOMP B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].  
 GN OMPB.  
 OS Rickettsia japonica.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 OX NCBI\_TaxID=35790;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=YH;  
 RA Uchiyama T.;  
 RT "Sequencing of the gene encoding the protein rOMP B of Rickettsia japonica."  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (By similarity).  
 CC -!- FUNCTION: THE 32 kDa BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.  
 CC -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPb FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB003681; BAA20138.1; --  
 DR InterPro; IPR006315; Autotransport.  
 DR InterPro; IPR005546; Autotransporter.  
 DR Pfam; PF03797; Autotransporter; 1.  
 DR TIGRFAMS; TIGR01414; autotrans\_bar1; 2.  
 DR Antigen; S-layer; Cell wall.  
 FT CHAIN 1 1338 120 kDa SURFACE-EXPOSED PROTEIN.  
 FT CHAIN 1339 1656 32 kDa BETA PEPTIDE.  
 FT DOMAIN 528 533 POLY-GLY.  
 FT SEQUENCE 1656 AA; 168097 MW; 3132A69C9DD5999F CRC64;  
 SQ

Query Match 6.1%; Score 157.5; DB 1; Length 1656;  
 Best Local Similarity 22.5%; Pred. No. 0.32;  
 Matches 136; Conservative 73; Mismatches 217; Indels 179; Gaps 35;

Qy 12 ASIAFGMAVSPVT---PIAFA-----AETGTVT-----QDTQKG-ATYKAVKVPDAE 55  
 STLVFDLA-NPTTQKAPLILADNALIVNGANGLVNTGFIQVSDKSPATVKAINIGDQG 240

Db 182 IDNAVSDSN-----KDGASYLIPQCKEAEYKASTDFNSLFTTTINGG 98

Qy 56 --IDNAVSDSN-----KDGASYLIPQCKEAEYKASTDFNSLFTTTINGG 98

Db 241 GFPMFNATNANALINQAGGTINFGTGTGRLLVSKNG---AATDFN--VTGSLGNG 295

Qy 99 RTVVTKKDTASANEIATWAKSISANTTPVSTVTSNNDGTE-----VINV----- 143

Db 296 LKGIIELTVAIN-----GQLIANAGPANAVICTNGAGRAAGFVSDVNGKAAITDQG 349

Qy 144 -----SQGYGVYSSTVNGAVIMVTSVTPNATIHEKNTDATWGDGGKTV 189

Db 350 VYAKDMVLOSANANGQVNFRIHVDVGIDGTTAFKTAASIVAITQNSFGTT--DFGNLAA 407

Qy 190 DOKTYSVGDITVKYTIITYKNAVNYHG--TEKYVQYVYKDTMPSASVVDLNEG-SYEVITTDG 247

Db 408 Q---VTVPDWTMTLGNFTGDANNPNTAGVITPAANGTCLASAS--ADANVAVTNNTAIEA 463

Qy 248 SGNITTLTGSEKATGCKYNNLENNNTTITPWAATVPTGNTONGANDDDFFYKGINITIT 307

Db 464 SG--GVGVQLSGTHTAELRLGNAGSVFKLA-----DGTVINGK-----VNG-T 503

Qy 308 VTYTGVLSKGA--KPGSADLPENTNIATINPNTSNDPFG--QKVTVRDQGITIKKIDGS 362

Db 504 VLVGVLAAGAITLDGS-----ATITGIDNGGGGAALQSITLAN-----DAT 546

Qy 363 TKASLQGAIFVLKNATGQFLNF-----NDTNNVEMWGEA 396

Db 547 KTLTLGGANII--SANGGTINFGANGGTTKLTSTONNIVVDCDLAIATDQTVVDASSLT 604

Qy 397 NATEYT-TGADGIITITGLKEGTYYLVEKAPL-GYNLLDNSQKVLGDGA----- 445

Db 605 NAQTLTISGTIGIIGANNNTTLGQFNIGSSKTTLNGGNVAIN--ELVTGNNGSVQFAHTY 662

Qy 446 --TDTTNS---DNLLVNPVTENN---KGTLPSTGG-IGTTFIYIIGA---ILVIGAG 491

Db 663 LITRTNAGQKKIIFNVVNNNTTLAAGTNLGSAAANPLAEINFGSGKARADTVLVNVEG 722

Qy 492 IVLVA 496

Db 723 VNLXA 727

RESULT 27  
 SLAP LACHE STANDARD; PRT; 439 AA.

ID SLAP LACHE STANDARD; PRT; 439 AA.  
 AC P38059;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE S-layer protein precursor (Surface layer protein).  
 GN SLPH.  
 OS Lactobacillus helveticus.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=1587;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CNRZ 892, and CNRZ 1269;  
 RA Callegari M.L., Cocconcelli P.S., Morelli L.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN 12)  
 RP SEQUENCE OF 31-53.  
 RC STRAIN=ATCC 12046;  
 RA Lortal S., van Heijenoort J., Gruber K., Sleytr U.B.;  
 RT "S-layer of Lactobacillus helveticus ATCC 12046: isolation, chemical characterization and re-formation after extraction with lithium chloride."  
 RT J. Gen. Microbiol. 138:611-618(1992).  
 RL J. Gen. Microbiol. 138:611-618(1992).  
 CC -!- FUNCTION: The S-layer is a paracrystalline mono-layered assembly of proteins which coat the surface of bacteria.  
 CC -!- SUBCELLULAR LOCATION: Cell wall. This bacterium is covered by a S-layer with hexagonal symmetry.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X91199; CA62606.1; --  
 DR EMBL; X92752; CA63409.1; --



Db 137 KKEPRQYPSFAESFNDNAYVLNRTSFGNGYYAGTWKSNKTSYTDATACLT-----GR- 190  
QY 101 YVTKDRTASA--NEATWAKSISANTTPVS-----TWTESNNDGTEVINVSQYGYVV 151  
Db 191 YATDPCYAGKLNIIITY--GLTKYDTPASGNAGGVITGNGGTG-----NTSNG----- 240  
QY 152 SSTVANGAVIMVTSVTPNATHEKNTDATMGDGGKTVDOKTSYSGDVTVKYITIKYNAVN 211  
Db 241 STSGNSGG-----SATTGTGTYTVKSGDSVMG-----ISHSFGITMAQLIEWNNIKN 287  
QY 212 ---YHGTKEVYQVINDKTPSASVVDLNEGSYEVTITDGSNITTLTQGE--KATG-KY- 265  
Db 288 NFIIYPG-----OKLTIKGGQ-SAGSSTTNGN--NASSGNTSGNTSGTQATGAKYT 339  
QY 266 -----NLLEN---NNF-----TTIPWAAVNTPTGNT- 290  
Db 340 VKSGDSVMKIANDHGISMNQLIEWNNIKNFYVPGQQLVSKGSSASGTSNTSTGNTS 399  
QY 291 QNGANDDFYKGINITVTVGLKSGAKPGSAD---LPENTNIATINPNTSNDDPGOK 346  
Db 400 SNTAN-----TGSTTSGSTYT--VRAGESVMSVSKFGISMNQLIOWNNIKNFYVPGOK 452  
QY 347 VTVRDGQITIKIDGTSKASLOCAIFVLKNATGQFLNFNDTNNVWEGTEANAT---EYTT 403  
Db 453 LIVKGG-----SSSNASTSTA-----NNKVTASNTSSATGQATYTV 491  
QY 404 GADGIITITGLKEGYVYLVEKAPGLYNLLDNS-----QKVLGDCATTTNSDNLVNP 458  
Db 492 KAGESVMVGANKNG-----ISMNQLIEWNNIKNFYVPGOKLIVKGGSSKA--SATATIKP 545  
QY 459 TVENNKGTLPSTGGIGITFIYI 481  
Db 546 TASTPAST--TPASSTGDTKYTV 567

## RESULT 29

BCA\_STRAG STANDARD; PRT; 1020 AA.  
AC Q02192;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE C protein alpha-antigen precursor.  
GN BCA.  
OS Streptococcus agalactiae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1311;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A909 / Serotype Ia;  
RX MEDLINE=93066179; PubMed=1438195;  
RA Michel J.L., Madoff L.C., Olson K., Kling D.E., Kasper D.L.,  
RA Ausubel F.M.;  
RT "Large, identical, tandem repeating units in the C protein alpha  
antigen gene, bca, of group B streptococci.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:10060-10064 (1992).  
CC -!- FUNCTION: MAY PLAY A ROLE IN BOTH VIRULENCE AND IMMUNITY.  
CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by  
an amide bond (Potential).  
CC -!- MISCELLANEOUS: IDENTICAL REPEATING UNITS DEFINE PROTECTIVE  
EPITOPES AND MAY PLAY A ROLE IN GENERATING PHENOTYPIC AND  
GENOTYPIC DIVERSITY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL; M97256; AAA26848.1; -.

DR InterPro; IPR005877; Gpos\_Ysirk.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF04650; Ysirk\_signal; 1.  
DR PRINTS; PR00015; GPOSANCHOR.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR TIGRFAMs; TIGR01168; Ysirk\_signal; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR Antigen; Cell wall; Peptidoglycan-anchor; Repeat; Signal.  
KW SIGNAL 1 41  
FT CHAIN 42 990 C PROTEIN ALPHA-ANTIGEN.  
FT PROPEP 991 1020 REMOVED BY SORTASE (POTENTIAL).  
FT DOMAIN 227 364 9 X 82 AA TANDEM REPEATS.  
FT REPEAT 965 976 INCOMPLETE.  
FT SITE 987 991 LPXTG SORTING SIGNAL (POTENTIAL).  
FT MOD RES 990 990 AMIDE-LINKED TO CELL WALL (POTENTIAL).  
SQ SEQUENCE 1020 AA; 108705 MW; 78035594FA190D5B CRC64;

## Query Match

Best Local Similarity 6.1%; Score 156.5; DB 1; Length 1020;

Matches 126; Conservative 64; Mismatches 236; Indels 151; Gaps 26;

QY 24 TPIAFAAETGTTTODTOKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQKAEYKA 83

Db 492 TPVSDKEITDLVKIPDGSKV-----PTVVGDRPDTPNPGDHKVTVEVTPDGTVDTEV 546

QY 84 ST-----DFNSLFTTTTNGGRTYVTKKOTASANEIATWAKSISANTTPVSTVT- 131

Db 547 TVHVTVPKVPDKKYDPT--GGETTVPGQTPVSDKEITDLVK-IPDGSKGVPTVVGDRP 602

QY 132 ESNNDGTEVINVSQYGYVVSSTVANGAVIMVTSVTPNATHEKNTDATMGDGGKTVDO 191

Db 603 DTNVEGDHKTVE-----VTYDPT--KDTVEVTVHVTVPKVPDKKYDPT---GGETTVPQ 654

QY 192 KT-----YSVGDTVKYITIKNA-----VNYHGTKEVYQVINDKTPSASVVDLNEG 238

Db 655 GTPVSDKEITDLVKIPDGSKGVPTVVGDRPDTPNPGDHKV---TVEVTP-----DGTKD 706

QY 239 SYEVTI-----TDGSGNITTLTQGEKATGKYNLLENNNFTITIPWAAVNTTP 286

Db 707 TVEVTVHVTVPKVPDKKYDPTGGETTVPGQTPVSDKEITDL-----VKIPDGSKGV 759

QY 287 TGNTQNGANDDFYKGINITVTVGLKSGAKPGSADLPENTNIATINPNTSND--DP- 343

Db 760 ---TVVGDRPDTPNPGDHKTVEVT-----YDPTGKDTVEVTVHVTVPKVPDKKYDPT 810

QY 344 GOKVTVRDG-----QIT--IKKIDGS-----TKA 365

Db 811 GGETTVPGQTPVSDKEITDLVKIPDGSKGVPTVVGDRPDTPNPGDHKVTVEVTPDGTKD 870

QY 366 SLQGAIFVLKNATGQFLNFNDT---NNVEWEGTEANATEYT-----TGADGIITITGLKE 416

Db 871 TVEVTVHVTVPKVPDKKYDPTGGETTVPGQTPVSDKEITDLVKIPDGSKGVPTVVGDRP 930

QY 417 GTYYLVEKAPGLYNLLDNSQKVLGDCATDTTN-----SDNLLNVTVE-----N 462

Db 931 DTNVEGDHKTVE-----EVTYDPTGKDTVEVTVHVTVPKVPDKKYDPTGKAQVN 981

QY 463 NKGTELPSTGGIGITFIYIIGAILVIGAGILVARRR 499

Db 982 GKGKLPATGENATPFFNVAALTISSVGLLSVSKKK 1018

## RESULT 30

APU THETU

ID\_APU THETU STANDARD; PRT; 1861 AA.

AC P38536;

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Amylopullulanase precursor (Alpha-amylase/pullulanase) (Pullulanase

DE type 1) [includes: Alpha-amylase (EC 3.2.1.1) (1,4-alpha-D-glucan

DE glucanohydrolase); Pullulanase (EC 3.2.1.41) (1,4-alpha-D-glucan





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:19:36 ; Search time 43 Seconds  
(without alignments)  
3683.492 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKMIQSLVSLASLAFGMVAV.....GAILVIGAGIVLVARRLRS 502

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database :

SPTREMBL.25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2578	100.0	502	2 Q84A41	Q84A41 streptococ
2	399	15.5	525	16 Q8G6G6	Q8G6G6 bifidobacte
3	331	12.8	625	16 Q836L8	Q836L8 enterococcu
4	324	12.6	522	16 Q8XP10	Q8XP10 clostridium
5	279.5	10.8	705	16 Q8DYR6	Q8DYR6 streptococ
6	265.5	10.3	535	2 Q68212	Q68212 actinomycet
7	265.5	10.3	665	16 Q97SC2	Q97SC2 streptococ
8	261	10.1	533	2 Q9X4D9	Q9X4D9 actinomycet
9	259	10.0	533	2 Q9X4D8	Q9X4D8 actinomycet
10	258.5	10.0	614	16 Q9CHH3	Q9CHH3 lactococcus
11	255	9.9	553	16 Q81D71	Q81D71 bacillus ce
12	251	9.7	507	2 Q48707	Q48707 lactobacill
13	250	9.7	533	2 Q9X4D2	Q9X4D2 actinomycet
14	250	9.7	705	16 Q88UC5	Q88UC5 lactobacill
15	244.5	9.5	554	16 Q85E64	Q85E64 streptococ
16	244.5	9.5	554	16 Q850S9	Q850S9 streptococ

17	243	9.4	535	2 Q9X4D1	Q9X4D1 actinomycet
18	232.5	9.0	674	16 Q8E4C4	Q8E4C4 streptococ
19	232	9.0	793	16 Q8YAG7	Q8YAG7 listeria mo
20	219	8.5	1612	16 Q8Y591	Q8Y591 listeria mo
21	217	8.4	562	16 Q8Y5H7	Q8Y5H7 listeria mo
22	216	8.4	1806	16 Q829J3	Q829J3 listeria in
23	215.5	8.4	3441	16 Q89PB9	Q89PB9 bradyrhizob
24	214.5	8.3	681	16 Q82FA4	Q82FA4 listeria in
25	212	8.2	2551	16 Q8CY18	Q8CY18 streptococ
26	208.5	8.1	476	16 Q836L9	Q836L9 enterococcu
27	208	8.1	917	16 Q88UJ0	Q88UJ0 lactobacill
28	206	8.0	520	16 Q8FMP6	Q8FMP6 corynebacte
29	205	8.0	1622	16 Q829J4	Q829J4 listeria in
30	204	7.9	3346	16 Q7WN54	Q7WN54 bordetella
31	200.5	7.8	827	16 Q8CPL9	Q8CPL9 staphylococ
32	200.5	7.8	893	16 Q97SC3	Q97SC3 streptococ
33	200	7.8	1441	16 Q9CFL1	Q9CFL1 lactococcus
34	199	7.7	1530	16 Q8Y479	Q8Y479 listeria mo
35	197.5	7.7	1006	16 Q88ZV1	Q88ZV1 lactobacill
36	195	7.6	2215	16 Q7WBN0	Q7WBN0 bordetella
37	194	7.5	1778	16 Q8FCB2	Q8FCB2 escherichia
38	193	7.5	454	16 Q8G453	Q8G453 bifidobacte
39	192.5	7.5	571	16 Q8YAG6	Q8YAG6 listeria mo
40	192	7.4	1744	16 Q82YW8	Q82YW8 enterococu
41	191	7.4	1711	16 Q8Y6L8	Q8Y6L8 listeria mo
42	191	7.4	2000	16 Q814R4	Q814R4 bacillus ce
43	189.5	7.4	5017	16 Q81FJ0	Q81FJ0 bacillus ce
44	189	7.3	723	16 Q8XP11	Q8XP11 clostridium
45	189	7.3	1519	2 Q48237	Q48237 helicobacte
46	188.5	7.3	628	2 Q9FD50	Q9FD50 streptococ
47	187	7.3	2402	2 Q9AER7	Q9AER7 staphylococ
48	186.5	7.2	1093	16 Q81ANG	Q81ANG bacillus ce
49	186.5	7.2	1275	5 Q76602	Q76602 caenorhabdi
50	186.5	7.2	1733	2 Q9K114	Q9K114 staphylococ
51	186	7.2	1036	2 Q8RM89	Q8RM89 streptococ
52	186	7.2	3705	16 Q8ZHA1	Q8ZHA1 yersinia pe
53	186	7.2	3710	16 Q8CZJ2	Q8CZJ2 yersinia pe
54	186	7.2	10791	16 Q7U7M8	Q7U7M8 synecococ
55	185.5	7.2	583	2 Q45822	Q45822 clostridium
56	185.5	7.2	1610	16 Q8FKP3	Q8FKP3 escherichia
57	185.5	7.2	1633	16 Q8CMP4	Q8CMP4 staphylococ
58	185	7.2	2768	16 Q8E9G6	Q8E9G6 shewanella
59	184.5	7.2	1330	2 Q934I7	Q934I7 microscilla
60	184.5	7.2	2468	16 Q912M3	Q912M3 pseudomonas
61	183.5	7.1	1183	16 Q8NUH0	Q8NUH0 staphylococ
62	182	7.1	775	2 Q9X6M6	Q9X6M6 enterococu
63	182	7.1	865	2 Q9FDJ1	Q9FDJ1 streptococ
64	181.5	7.0	1315	2 Q8E488	Q8E488 staphylococ
65	181.5	7.0	1371	16 Q8XQ42	Q8XQ42 ralstonia s
66	181	7.0	1129	16 Q8XL11	Q8XL11 clostridium
67	181	7.0	3029	16 Q55582	Q55582 synecocyst
68	181	7.0	5010	16 Q81CV2	Q81CV2 bacillus ce
69	180.5	7.0	3930	16 Q98E20	Q98E20 rhizobium l
70	180	7.0	1166	2 Q8E489	Q8E489 staphylococ
71	180	7.0	1231	2 P72362	P72362 streptococ
72	180	7.0	1389	16 Q8E1C4	Q8E1C4 streptococ
73	180	7.0	2353	2 P71401	P71401 haemophilus
74	180	7.0	5017	16 Q81SN0	Q81SN0 bacillus an
75	179.5	7.0	1260	2 Q9XDB6	Q9XDB6 streptococ
76	179	6.9	1141	16 Q8NXX5	Q8NXX5 staphylococ
77	179	6.9	1307	16 Q8EV70	Q8EV70 mycoplasma
78	178.5	6.9	4106	16 Q8XQP2	Q8XQP2 ralstonia s
79	178.5	6.9	1614	2 Q9KKB6	Q9KKB6 astrakhan r
80	178.5	6.9	2566	17 Q8TSE7	Q8TSE7 methanogarc
81	178.5	6.9	4336	16 Q8YKJ3	Q8YKJ3 anabaena sp
82	178	6.9	1195	3 Q96WU8	Q96WU8 schizosacch
83	178	6.9	3705	2 Q9F285	Q9F285 yersinia pe
84	177.5	6.9	1268	2 Q9RER7	Q9RER7 bacillus ce
85	177.5	6.9	1965	16 P73089	P73089 synecocyst
86	177	6.9	440	2 Q07120	Q07120 lactobacill
87	177	6.9	1029	2 Q52708	Q52708 rickettsia
88	177	6.9	1347	16 Q8NXX6	Q8NXX6 staphylococ
89	176.5	6.8	1385	16 Q99W47	Q99W47 staphylococ

90	176.5	6.8	1794	16	Q8RHH1	Q8rhh1 fusobacteri
ALIGNMENTS						
RESULT 1						
Q84A41 PRELIMINARY; PRT; 502 AA.						
ID	Q84A41					
AC	Q84A41					
DT	01-JUN-2003 (TrEMBLrel. 24, Created)					
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)					
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)					
DE	Spb1.					
DE	Streptococcus agalactiae.					
GN	SPB1.					
OS	Streptococcus agalactiae.					
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;					
OC	Streptococcus.					
OX	NCBI_TaxID=1311;					
RN	[1]					
RC	SEQUENCE FROM N.A.					
RC	STRAIN=874391;					
RA	Aderson E.E., Bohnsack J.F.;					
RT	"Subtractive Hybridization Identifies a Novel Adhesin/Invasin of					
RT	Virulent Serotype III Group B Streptococcus agalactiae.";					
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; AF485279; AA049409.1; -					
DR	InterPro; IPR001899; Gram_pos_anchor.					
DR	Pfam; PF00746; Gram_pos_anchor; 1.					
DR	TIGRFAMS; TIGR01167; LPXTG anchor; 1.					
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.					
SQ	SEQUENCE 502 AA; 53450 MW; E3B7F5603F6D192 CRC64;					
Query Match 100.0%; Score 2578; DB 2; Length 502;						
Best Local Similarity 100.0%; Pred. No. 2.5e-109;						
Matches 502; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1	MKKMIQSLVASLAFGMVSPVTPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNAN	60			
DB	1	MKKMIQSLVASLAFGMVSPVTPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNAN	60			
QY	61	VSDSNKDGASYLIPQGEAEYKASTDFNSLFTTTTNGGRTYVTKOTASANEIATWAKSI	120			
DB	61	VSDSNKDGASYLIPQGEAEYKASTDFNSLFTTTTNGGRTYVTKOTASANEIATWAKSI	120			
QY	121	SANTTPVSTVTSNNDGTEVINVSQYGYVSVTVNNGAVIMVTSVTPNATIEHKNTDAT	180			
DB	121	SANTTPVSTVTSNNDGTEVINVSQYGYVSVTVNNGAVIMVTSVTPNATIEHKNTDAT	180			
QY	181	WGDGGKTVDPQKTVSGDVTYKITYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSY	240			
DB	181	WGDGGKTVDPQKTVSGDVTYKITYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSY	240			
QY	241	EVITIDSGNITLTQGESEKATKYNLLENNFTIIPWAAATPTGNTONGANDDDFFY	300			
DB	241	EVITIDSGNITLTQGESEKATKYNLLENNFTIIPWAAATPTGNTONGANDDDFFY	300			
QY	301	KGINTIVTVTVGLKSGAKPGSADLPENTNIATINPNTSNDPQGVTVRDGQITIKKID	360			
DB	301	KGINTIVTVTVGLKSGAKPGSADLPENTNIATINPNTSNDPQGVTVRDGQITIKKID	360			
QY	361	GSTKASLQGAIFVLKNATQFLNPDNNVWEGTEANATEYTTGADGIIITGLKEGTY	420			
DB	361	GSTKASLQGAIFVLKNATQFLNPDNNVWEGTEANATEYTTGADGIIITGLKEGTY	420			
QY	421	LVEKAPLGNLNDNSQKVLGDGATDTTNSDNLNVNPTVENNKGTELPSTGGITTFY	480			
DB	421	LVEKAPLGNLNDNSQKVLGDGATDTTNSDNLNVNPTVENNKGTELPSTGGITTFY	480			
QY	481	IIGAILVIGAGIVLVARRRLS 502				
DB	481	IIGAILVIGAGIVLVARRRLS 502				
RESULT 2						
Q8G6G6 PRELIMINARY; PRT; 525 AA.						
ID	Q8G6G6					
AC	Q8G6G6					
DT	01-MAR-2003 (TrEMBLrel. 23, Created)					
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)					
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)					
DE	Possible cell surface protein similar to FimA fimbrial subunit of					
DE	Actinomyces naeslundii.					
GN	BL0675.					
OS	Bifidobacterium longum.					
OC	Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;					
OC	Bifidobacteriaceae; Bifidobacterium.					
OX	NCBI_TaxID=216816;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=NCC 2705;					
RC	MEDLINE=22294977; PubMed=12381787;					
RA	Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,					
RA	Pessi G., Zwhilen M.-C., Desiere F., Bork P., Delley M.,					
RA	Pridmore R.D., Arigoni F.;					
RT	"The genome sequence of Bifidobacterium longum reflects its adaptation					
RT	to the human gastrointestinal tract.";					
RL	Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).					
DR	EMBL; AE014689; AAN24496.1; -					
DR	InterPro; IPR008454; Cna_B.					
DR	InterPro; IPR001899; Gram_pos_anchor.					
DR	Pfam; PF05738; Cna_B; 1.					
DR	Pfam; PF00746; Gram_pos_anchor; 1.					
DR	TIGRFAMS; TIGR01167; LPXTG anchor; 1.					
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.					
KW	Complete proteome.					
SQ	SEQUENCE 525 AA; 55559 MW; 11D954D398C55E75 CRC64;					
Query Match 15.5%; Score 399; DB 16; Length 525;						
Best Local Similarity 27.3%; Pred. No. 1.1e-10;						
Matches 160; Conservative 72; Mismatches 198; Indels 156; Gaps 26;						
QY	3	KKMIQSLVASLAFGMVSPVTPIAFAAETGTTVQDTQ---KGATYKAYKVFDAEIDNA	59			
DB	6	RKVAAGLAAATMLGATGATTASAEATGTLTSSDAAFNGKKVKNAYQMFSAFDAA	65			
QY	60	NVSDSNKDGASYLIPQGEAEYKASTDFNSLFTTTTNG-GRTYVTKKDTAS-----	109			
DB	66	-----GQATYTLASPWNFFKTNADLGITGVTDA NVSEKAYDYVYSL	109			
QY	110	-----ANEIATWAKSI SANTTPVSTVTSNNDGTEVINVSQY-----GYYY	150			
DB	110	NQDMKLABFATKASNWAKKAGNNVSAVNPAAATKASG-----SQYVATFANLPLGYV	163			
QY	151	VS-----STVN-----NGAVIMVTSVTPNATIH-----EKN TDA-TWGDG---CGKTV	192			
DB	164	VSPEAGSTNKRHTDAMLNVNVTATTNINLKSEYPTVVDKTDADKKGDSAQIGSKVNFQL	223			
QY	193	TVSVGTVTKYTYTYKNAVNYHGTEKYQYVVKDTMPSASVVDLNEGSYEVTITDGSNIT	252			
DB	224	KSKVPTSEY-----NYVFKIVDTLSAG--LDENN---DVTVKVGDTALT	264			
QY	253	TLTQGESEKATKYNLLENNFTIIPWAAATPTGNTONGANDDDFFYKGINITVITYTG	312			
DB	265	ATTDSVTTTKG-----TVTI-----DLSNVYKTDNASKAGKGIIVTYS	304			
QY	313	VLKSGAKPGSADLPENTNIATIPNTSNDPQGVTVRDGQIT-----IKKI--DGS	362			
DB	305	TLNENAFVGTGPDQNNPGLNLSAKVQYNGPSEENIGESTPSETHSYTFNFKKIKYKEGD	364			
QY	363	TKASLQGAIFVLKNATQFLNPDNNVWEGTEANATEYTTGADGII	409			
DB	365	TENALAGAKFQLLDSDKTVISLVKKSNDVYRPAKTSDDDEV-----TEVETPATGII	416			
QY	410	TITGLKEGTYVLEKKAPLGNLNDNSQKVLGDGATDTTNS-DNLLVN---PT-----	459			

```
Db 417 EFTGLKAGTYLKEFPAPKYNKLSDPVKVTINATINKTTGALESWTNVSAPTADVTVP 476
QY 460 ---VENNKOTELPSTGGIGTITFIYIGAILV-IGAGIVLVARRR 501
Db 477 VVKIENKKGALLPDTGGMGTVLFTFGVLIVAGAGYVYKSNRKS 522

RESULT 3
Q836L8
ID Q836L8 PRELIMINARY; PRT; 625 AA.
AC Q836L8;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN EF1093.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=2250857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
RA Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
RA Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
RA Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis."
RL Science 299:2071-2074(2003).
DR EMBL; AF016950; AAO80893.1; -.
DR TIGR; EF1093; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAM; TIGR01167; LPXG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Complete proteome.
SQ SEQUENCE 625 AA; 67989 MW; EB97450BB11F0DE CRC64;

Query Match 12.8%; Score 331; DB 16; Length 625;
Best Local Similarity 26.1%; Pred. No. 1.6e-07;
Matches 148; Conservative 78; Mismatches 227; Indels 114; Gaps 26;

QY 3 KKMIOQLIVASLAFGNVSPVPIAFAA--ETGTTTVDQTKGATYKAYKVFDAIDNAN 60
Db 100 KQAVQSL-----TPGKPVAGTTDANGNTVQLPKQNGKDAVYTIKEPKEGV 148

QY 61 VDSNKGDA--SYLIPQKGEABYKASTD-----FNSLFTTTTNG--G 98
Db 149 VAATNMVAFPPYEMIKOTDGSYKYGTEBLAVVHVIPKMNVDGSLKKVGTAEENGLNG 208

QY 99 RYVTKDPTASANEIA-----TWA-----KSIANTTPV--STVTESNNDGTE 139
Db 209 AEFVLSKSESGPGTKYIOGVKDGXYLTWTDKEQAKRFITGKSYSEIGNDFTEAEN--GTG 267

QY 140 VNVS--QYGYTVVS--STVNVGAVIMTVSVP-----NATIEK--NTDATWGGGKG 187
Db 268 ELTVKNLEGVSVILEVKAPNNAELIENQTKTPFTEIANNQTPVEKTVKNDTSKVDKTP 327

QY 188 TVDQKTSYVGDVTVKYITTYKNAVNYHGTE---KVYQVVIKDTMPSASVVD--LNESGYEV 242
Db 328 SLDGKDVAIKEIKYQISVNIPLGIADKEGDANKYKFNVLVDKDAALTFDNTVSGEYAY 387

QY 243 TITDGSNITLTITQSEKATGKNLLENNFTIYI--PWAATNPTGNTGANDDDFF 300
Db 388 ALYDGDVTI-----APENYQVTEQANGFTVAVNPAYIPTLTPGG----- 426

QY 301 KGINTITVTYTVGLSKAGPGSADLPENTNIATINPNTSNDPQGVTVRVDQGITIKKID 360
Db 427 ----TLKFVYFHLNEKADPTKG-----FKNEANVDNGHTDDQPTPTVEVTGGKRFIKVD 478
```

```
QY 361 GSTKA--SLOGAIFVLKNA---TQFLNPNDTNV--VIEW--GTANATEYTTGADGIITING 413
Db 479 GDVTAQALAGASFVVRQNSDTANYLKIDETTKAATWVKTAEBATFTTTADGLVDITG 538
QY 414 LKEGYTYLVEKKAPLGVNLLDNSQKVLIGDGDATDTTNSNLLVNPVTEN--NGKTELPSTG 472
Db 539 LKYGTYYLEETVAPDDYVLLTNRIEFVNVQESYGTT--ENLVSPEKVPKHKGT--LPSTG 595

QY 473 GIGTTFIYIIGAILVIGAGIVLVARR 499
Db 596 GKGIYVILGSGAVLLIAGVYFARRK 622

RESULT 4
Q8XP10
ID Q8XP10 PRELIMINARY; PRT; 522 AA.
AC Q8XP10;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Probable surface protein.
GN CPE0156.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX MEDLINE=21664373; PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP003185; BAB79862.1; -.
DR InterPro; IPR008454; Cna_B.
DR Pfam; PF05738; Cna_B; 1.
KW Complete proteome.
SQ SEQUENCE 522 AA; 57247 MW; 8076613B81C314EA CRC64;

Query Match 12.6%; Score 324; DB 16; Length 522;
Best Local Similarity 25.7%; Pred. No. 2.7e-07;
Matches 143; Conservative 66; Mismatches 187; Indels 160; Gaps 27;

QY 19 AVSPVPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASYLIPQKGE 78
Db 32 SISKDAPL-----KGSIII--SKKGATFTAYKLDA-----IKSGDA 66

QY 79 ABYKASTDPNSLFTTTTNG--GRYVTVTKDTASANEIATWAKSISANTTPVSTVTESNND 136
Db 67 YEYSVNSDLKOFFNNSNYGSYSQESIQKLNGBQVKEFA-----INLH----KVILENKS 117

QY 137 GTEV-----INVSYQYVYVSVSTVNN--GAVIMVT-----SVTP--NA 170
Db 118 GOELKDGQKNTVDLGYLVLTETSDSDEGAASVASTPIIVSPQVSGDSMNYDITNPKDNT 177

QY 171 TIHEKNVDATWGGGKTVQKTSYVGDVTVKYITIT-----YKNAVNYHGTEKVKYQYV 224
Db 178 PILEKNIVKE-----NQRVKTSSSENIQDVVKYEVKASIPVYQKNAQN-----IMYKFT 225

QY 225 DTMPSASVVDLNEGSYEVITIDG--SGNITTLTQGSSEKATGKNLLENNFTIIPWAA 283
Db 236 DTMSKGLTYDEKGT--FKVTSGDKVPAKDTDYTDVVK-----QEDGSTVITI----- 271

QY 284 NPTGNTGANDDDFFYKGINF-----ITVYTVGLVKSAGKPGSADLPENTNIA-----TI 334
Db 272 -----NFVYENIKAYAEITGLNYQATLNKDAVISNKENLGNLTNNIQLDYN 318

QY 335 NPNT--SNDPQGVTVRVDQGITIKKIDGSTKAS--LOGAIFVLKNAATGOFL---NFNDIN 388
Db 319 NPHVKDSYKLTDRKVTITVTFGGITKVDSELSNKLQGAEFVSKDAGGKIVAKYTYDEKG 378
```

```

QY 389 NVEWTEANATEYTTGADGIITITGLKEGTYLVLEKKAFLGVNLLDNSOKVIL----- 441
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
379 QVYV-LSNG----VTNSKGITITPLGLKEGKYFITEEVAFSGYSLKKNPVEVITANKDES 434
QY 442 -----GQADTTNSD-NLLNVPVENKKGTELPSTGGIGTTTPIYG 483
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 GNYTGAATIEISNGKAGQIINDISEKDGNIILFNVOIENHAGFSLPSTGGLNGTGFIKIA 494
QY 484 AIL-----VIGAG 491
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 IILLSIVCLAILGLG 510

RESULT 5
Q8DYR6 PRELIMINARY; PRT; 705 AA.
AC Q8DYR6;
DT 01-MAR-2003 (T-REMBLrel. 23, Created)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN SAG1407.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547;
RA Tettelin H., Masiugi V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014256; AAN00278.1; -
DR TIGR; SAG1407; -
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMS; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Complete proteome.
SQ SEQUENCE 705 AA; 75475 MW; 7BB1774ECBBE1A1A CRC64;

Query Match 10.8%; Score 279.5; DB 16; Length 705;
Best Local Similarity 24.3%; Pred. No. 3.9e-05;
Matches 149; Conservative 76; Mismatches 213; Indels 175; Gaps 30;

QY 19 AVSPVTP-----TAFRAE---TGTTVODTQKGATY-----KAYKVFDAEIDNANV 61
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 AVSGLTPEGTGLVNTKGLKGEFKIVEVSKSYNNNGSLLAASKAVPV-----NITL 195
QY 62 SDSNKGDA---SYLIPQKREA-----EYKASTDFNSLFTTTTNGRTYVTKD 106
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 PLVNEGDGVADAHVYPKNTKEPKIDKNFAKTNLTALTVDNRLTAGANYG-NYARDKA 244
QY 107 TASN-----EIAI-----WAKSISANTPTVSTTESNNDTEV----- 140
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 TATAEIGKVVPYEVKTKIHKRSKYENLVWTDIMSNGLTMSGTVSLKASGTTTFAXD 304
QY 141 -INVSQGY-----YVVSSTVNNGAVIMVTSVTENATHEK 175
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 ELSIDARGFTLKFTADGLGLEKAAKTAADIEFLTISATVNGQAIL-----DNPSN 356

```

```

QY 176 NTDATWGGGGKTVQD--KTVSVGDTVKYITTYKNVNVHTEKYQYVIKD----- 225
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
357 DIKLSYGNKPGKDLTPELPVTPSKGE-VTVAKTWSGDIAPDGVNVVYTLKDKDKTVAASVL 415
QY 226 TWPSASVDLNEG-SYEVTITDGSNIITLTGQ-----SEKATGKNLLE-ENNFT 275
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 TKTSKGTIDLNGIKFEV-----SGNFSKFTGLENKSYMISERVSGYGSAINLENGKVT 470
QY 276 ITIPWAATN-TPTGNTQ-----NGANDDFYKGINITITVITYTG---VLKSGA 318
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
471 INTKDSNDPTPLNTEPEKVEHGHKKFVKNEQGBL--AGAQFVVKNAGKYLALKADQ 528
QY 319 KPGSADLPE-----NTNIATINPNTSDPDQKQVTRDQIITIKKIDGSTRASLOCAIV 373
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
529 SEGQKTLAAKIALDEALAAVYNKLSATDQKGEK-----GITAKELIKTKQADYDAAFTE 582
QY 374 LKNATGQFLNFDNTNVEGTE-ANATEYTTGADGIITITGLKEGTYLVLEKKAFLGVN 432
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
583 AETA-----YEWITDKARAITYSNDQGFVETGLADGTYNLEETLAPAGFAK 630
QY 433 LDNSOKVILGCGATDTTNSDNLVNPT-----VENNKGTELPSTGGIGTTTIFYIIGAIL 486
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
631 LAGNIKFFVNOGSYITGGNIDVANSNQKDATRVENKVT-IPQTGGIGTILFTIIGLSI 689
QY 487 VIGAGIVLVARRR 499
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
690 MLGA-VVIMKRQ 701

RESULT 6
ID 068212 PRELIMINARY; PRT; 535 AA.
AC 068212;
DT 01-AUG-1998 (T-REMBLrel. 07, Created)
DT 01-AUG-1998 (T-REMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (T-REMBLrel. 25, Last annotation update)
DE Fimbrial structural subunit.
GN FIMA.
OS Actinomycetes naeslundii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=T14V;
RX MEDLINE=98187920; PubMed=9529071;
RA Yeung M.K., Donkersloot J.A., Cisar J.O., Ragsdale P.A.;
RT "Identification of a gene involved in assembly of Actinomycetes
RT naeslundii T14V type 2 fimbriae.";
RL Infect. Immun. 66:1482-1491(1998).
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY
CC AN AMIDE BOND (BY SIMILARITY).
CC EMBL; AF019629; AAC13545.1; -
CC GO; GO:0005618; C:cell wall; IEA.
CC InterPro; IPR008454; Cna_B.
CC InterPro; IPR001899; Gram_pos_anchor.
CC Pfam; PF05738; Cna_B; 2.
CC Pfam; PF00746; Gram_pos_anchor; 1.
CC TIGRFAMS; TIGR01167; LPXTG anchor; 1.
CC PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Cell wall; Peptidoglycan-anchor.
SQ SEQUENCE 535 AA; 56038 MW; B115867F500C1356 CRC64;

Query Match 10.3%; Score 265.5; DB 2; Length 535;
Best Local Similarity 26.8%; Pred. No. 0.00012;
Matches 151; Conservative 58; Mismatches 228; Indels 127; Gaps 29;

QY 11 VASLAP-GMAVSPVTPPIAFAET---GTTVQDTQKGATYKAYKVFDAEIDNANVSDSNK 66
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
18 VLTALFLGLA-----PSAVATETPNYGNIKTKDATGSLAIHKHLNGGGKIGNFTGTPQNA 72
QY 67 DGASYLIPQKAEAYKASTDFN---SLFTTTTNGRTYVTKDTSANAEIATWAKSISA 122

```

```

Db 73 DSKGTPV-QGVVFTAYPITDNLKDPAGWDTISLDSKAGV--PDSACTNPAAPTGLCAHKE 129
Qy 123 NTTPVSTVSTESNDNGTEVINVSQGYVYSSVTNNGAVI-----MWTSTVTPNATHIERN 176
Db 130 GTPKVSPT--NDEGLATITEMPQVQYALVCEITTPGNIQKAKPFVTVTPHNTA--AG 184
Qy 177 TDATW-----GDGCKTVDOKT---YSVGDPTVKYTIYKNAVNHGTEKYVQYV 222
Db 185 ADGOWIYDVHVPKNEAISVEKSIQCKLNGYGVSLIKFPVS--STAPTLDKSFYKYP 242
Qy 223 -IKDTPSASVVDLNEGSYEVTITDGSNITITLQSGSEKATGKYNILLENNTTITIPWA 281
Db 243 QLRDT-----LDDRLTAVTATEVSLEGTITLDPDYKV-----DTKGTVTVTFT 286
Qy 282 ATNPTFGNTQNGANDFFYKGINIITVTVGLKSGAK-----PGSADLPENTN 330
Db 287 AEGLKIKKAAPGKVSAPVQ--KVTPEARNGAITNRAQVISDVTVAEQPTPEEPANPE 344
Qy 331 IATINPNTSNDPQKVTVRDGOITIKKID---GSTKASLOGAIIVLKNATCOFLNFD 386
Db 345 ----NPTTSNE-----VTSRWGDLKKVDNHQOQDKAGLOAQPOLYKAKNAVAG-TC 394
Qy 387 TNNVWGTGA--NATEYTTGADGIITITGL-----KEGT---YVLVEKKAP 427
Db 395 TKDKEGDPAINGETTLTTDAQAINVKGLFISDSIDGANRDNQKDATARCIVLVETKAP 454
Qy 428 LGHNLNDSQKVLGDGA-----TDTNSNMLNVPVNNKGT--ELPSTGGIGTTI 478
Db 455 AGTVL-----PAGDGAVTPVKIEGAVTTDNV-----TIENTKQSVPGPLTGANGMLI 503
Qy 479 FYIIGA-ILVIGAGIVLVARRLR 501
Db 504 LTASGASLLMIAVGSVLVARYRER 527

RESULT 7
Q97SC2 PRELIMINARY; PRT; 665 AA.
ID Q97SC2 AC Q97SC2;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN SP0463.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.B., Lewis M.R., Radune D.,
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Lofus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AB007358; AAK74623.1; -.
DR FIC; F95053; F95053.
DR TIGR; SP0463; -.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B_1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXPG_anchor; 1.
KW Complete proteome.
SQ SEQUENCE 665 AA; 71364 MW; 6F2BD7616398CBA5 CRC64;

```

```

Query Match 10.3%; Score 265.5; DB 16; Length 665;
Best Local Similarity 24.4%; Pred. No. 0.00016;
Matches 164; Conservative 63; Mismatches 227; Indels 219; Gaps 33;

Qy 9 LLVASLAFGMVSPVPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDANVSDSNKDG 68
Db 15 LUTASLSFAAT-----VFAAGTTTSTVTHKLATDGMKDIANELETGNTA-GNKVG 67
Qy 69 ASYLPQGGKEABYKASTDFNSLFTTTN-----GGRYVTVTKDITASANBIATWAKISAN 123
Db 68 ---VLP-----ANAKEIAGVMFVWTNNIEIDENGQTLGVNIDP-----QIFKLSGAMP 114
Qy 124 TTPVSTVSTESNDNGTEVIN--SQGYVYVSS-----TVNNGAVIMVTSVTP 168
Db 115 ATAMKLTAEAGKNTANLPAKYKIYHLSLTVYVGEDGATLTGSKAVPIEIELPLND 174
Qy 169 --NATHEKNTDA---TWGDDGGKT-----VDQKT---YSVGDPTVKYTIYK--NAVNY 212
Db 175 VVDAHVPKNTAKPKIDKDFGKANPDPTRVDKOTPVNHHQGVGVVEYIVTKIPALANY 234
Qy 213 -----HGTEKYV-----QYVIKDT-----MPSASVVDLNEG-- 238
Db 235 ATANWSDRMTEGLAFNKGITVTVDDVALEAGDYALTEVATGFDLKLTDAGLAKVNDQNA 294
Qy 239 -----SYEVTITDGS-----GNITTLQSGSEKATG---KYNLLENNNTTITIPWA-A 282
Db 295 EKTVKITYSATLNDKAIIVEPESNDVTFNYPNNPDHNTPKPNKPNENGDLTLTKTWVDA 354
Qy 283 TWTPTG-----NTQNGANDFFYKGINI-----TVTYTGVLSK-----G 317
Db 355 TGAPIPAGABATFDLVNAQTG-----KVQTVTLTDDKNTVTVNGLDKNTYEKFEVER 407
Qy 318 AKPGSADLPENTNIAT-----NPNTSNDPQKVTVRDGOITIKKIDGSKASIQGAI 371
Db 408 IKGYSADYQEIITAGELAVKWKDENPKPLDTEPKVTVYTKKFKVNDKDNK--LAGAE 465
Qy 372 FVLKNA--TGQFL-----KVVQTVTLTDDKNTVTVNGLDKNTYEKFEVER 382
Db 466 FVIANADNAGQYLARKADKVSQEEKVLVTTKDALDRAVAAYNALTAQOQTOEKEKVDK 525
Qy 383 ---NEN-----DTNNVWGTG---ANATEYTTGADGIITITGLKSGTYVYLVEKKAELGVN 431
Db 526 AQAAVNAVIAANAFEFWADKNENNVKLVSDAQGRFEITGLLAGTYLEETKQPAGYA 585
Qy 432 LLDNSQKV-----ILGDGATDTTNSDNLNVPVNNKGTETLPSTGGIGTTFIIG 483
Db 586 LLTSRQKEVTSATSYSGGIEYTAGSK--DDATKVVNKKITIPQGGIGTIIPAVAG 643
Qy 484 AILVIGAGIVLVA 496
Db 644 AAAM--GIAVVA 653

RESULT 8
Q9X4D9 PRELIMINARY; PRT; 533 AA.
ID Q9X4D9 AC Q9X4D9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type-1 fimbrial major subunit precursor.
GN FIMP.
OS Actinomyces naeslundii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycineae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LY7;
RX MEDLINE=99242779; PubMed=10225854;
RA Li T., Johansson I., Hay D.I., Stromberg N.;
RT "Strains of Actinomyces naeslundii and Actinomyces viscosus exhibit

```

```

RT structurally variant fimbrial subunit proteins and bind to different
RT peptide motifs in salivary proteins."
RL Infect. Immun. 67:2053-2059(1999).
DR EMBL: AF107020; AAD28829.1; -.
DR InterPro: IPR008454; Cna_B.
DR DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00738; Cna_B; 2.
DR DR Pfam: PF00746; Gram_pos_anchor; 1.
DR DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 30 POTENTIAL.
SQ SEQUENCE 533 AA; 57062 MW; F247CCACED4E3F38 CRC64;

Query Match 10.1%; Score 261; DB 2; Length 533;
Best Local Similarity 25.1%; Pred. No. 0.00019;
Matches 146; Conservative 60; Mismatches 211; Indels 164; Gaps 29;

QY 13 SLAFGMVSPVTPPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSN--KDGAS 70
Db 17 TLAAGALVSP-TGAAPVADPNGSTI-DPQAATLTIVHK-----CEQDTNGVKEGAG 66
QY 71 YLTPQGEAEYKASTDFNSLFT-----TTTNGGRT-----YVTKDITASANEIATWAK 118
Db 67 NEDPQ---AECKPVDVEFTITKLNVDLTVDGKTLADFKGVVVRAGALK-----STTVQ 119
QY 119 SISANTTPVSTVTSNNDGTEVINVSQYGYVVSSTVNNGAVIM-----V 163
Db 120 KITTGANGLASFTDAQ---TEV-----GAYLSETRTPDKVIPAEDFVVTLPMTPQDT 170
QY 164 TSVTPNATIEKNTDNTWGGGKTV--DQKTVSGVTVKYTIYK-NANNYHGTETKVQY 221
Db 171 TKWYNVHVYPKNTLS-----GVDKQVTDKPAKSGRDIYTIITTSIPKVDYPGGARIKRY 226
QY 222 VIKDTM-----PSASVVDLNEGSYEVTITDGSNITTLTQSEKATGKYNLLE- 270
Db 227 EVVDRLDKRIKKEALTPVVKIV---GQNEVTIADTDTYTLTAEKGDHNWATIQLTBEG 282
QY 271 -----NNN-----FTTIPWAATNPTGNTONGANDDFYKGINTIITVYTGVLK 315
Db 283 RRKASEARYNGETKQVLTAKFDAAVNLEGLSNTA----- 321
QY 316 SGAKPSADLPENTNATINPTNSDDPG---QKVTVRDQGIITIKIDG---STKASLOG 369
Db 322 -GLIPN--DSP-NFTWDPNNPGTTDIPGIPPTPVLISKYKGVLTITGTDQLADTKYNG 377
QY 370 AIF-----VLKNATGQFLNEND--TNNVWGTEANATEYTTGADGIITITGLK----- 415
Db 378 AQFQVECTKTAGAMLRDSDPTQVDPITIGKEKFTTAGOGTVEINVLRANDYVNGA 437
QY 416 -----EGTYLVEKAPILGYN-----LLDNSQKVLGDGATDTTNSDNLVNP 458
Db 438 KKQLTDEDDYVCLVETKAPESYLOADPLPFRVLAKEAEK---KAATEVTITD----- 487
QY 459 TVENNKGTETLPSTGGITTFYIIGALLVIGAGILVARRR 499
Db 488 -IPKNAGFRPLTGANGVIFLTAGALLVAGGAVVAYANKR 527

RESULT 9
Q9X4D8 PRELIMINARY; PRT; 533 AA.
AC Q9X4D8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Type-1 fimbrial major subunit precursor.
GN FIMP.
OS Actinomyces naeslundii.
OC Bacteria; Actinobacteridae; Actinomycetales;
OC Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1655;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=P-1-K;
RX MEDLINE=9242779; PubMed=10225854;
RA Li T., Johansson I., Hay D.I., Stromberg N.;
RT "Strains of Actinomyces naeslundii and Actinomyces viscosus exhibit
RT structurally variant fimbrial subunit proteins and bind to different
RT peptide motifs in salivary proteins."
RL Infect. Immun. 67:2053-2059(1999).
DR EMBL: AF107019; AAD28828.1; -.
DR InterPro: IPR008454; Cna_B.
DR DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00738; Cna_B; 2.
DR DR Pfam: PF00746; Gram_pos_anchor; 1.
DR DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 1 30 POTENTIAL.
SQ SEQUENCE 533 AA; 56772 MW; 6261AC963AD1D75C CRC64;

Query Match 10.0%; Score 259; DB 2; Length 533;
Best Local Similarity 26.1%; Pred. No. 0.00024;
Matches 148; Conservative 67; Mismatches 216; Indels 136; Gaps 31;

QY 13 SLAFGMVSPVTPPIAFAAETGTTTVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASYL 72
Db 17 TLAAGALVAP-TGAAPADPNGSTI-DPDAATLTIVHKCEQDTNGVKEGTGNE----- 69
QY 73 IFQGEAEYKASTDFNSLFT-----TTTNGGRTV-TKKDTASANEIATWAKSISANTT 125
Db 70 -PQ---AECKPVDVEFTITKLNVDLTVDGKTLADLKGVDVVKAG--ALKSDTVQKIIT 123
QY 126 PVSVTESNNDGTEVINVSQYGYVVSSTVNNGAVI-----MVT--SVTP-----NA 170
Db 124 GANGLASFTDAQTEV-----GAYLSETRTPDKVIPAEDFVVTLPMTPQDTAKWYNV 177
QY 171 TTHEKNTDNTWGGGKTV--DQKTVSGVTVKYTIYK-NANNYHGTETKVQYVIKDTM- 227
Db 178 HVPKNTLS-----GVDKQVTDKPAKSGRDIYTIITTSIPKVDYPGGARIKRYEVDRLD 233
QY 228 -----PSASVVDLNEGSYEVTITDGSNITTLTQSEKATGKYNLLENNFTITI 278
Db 234 KRIKKEALTPVVKIV---GQNEATLANCT-DYTLIT-----AEK-----DHNWATIQL 278
QY 279 PWAATNPTGNTONGANDDFYKGINTIITVYTGVLKS-----GAKPGSAD-LPEWTNIA 332
Db 279 TEEGRKSEARYNGND-----TKIEVTLTAKLDAAVNLEGLSNTAGFIPNDSPNF 331
QY 333 TINPN---TSNDDPG---QKVTVRDQGIITIKI---DGSTKASLOCAIFVL---KNATG 379
Db 332 TWDPNPGTTDIPGIPPTPVLISKYKGVLTITGTDLADTKYNGAQFQVECTKTASG 391
QY 380 QFLNEND--TNNVWGTEANATEYTTGADGIITITGLK-----EGTYLVL 422
Db 392 ATLRLSDPTQVDPITIGKEKFTTAGOGTVEINVLRANDYVNGAKKQLTDEDDYVCLV 451
QY 423 EKKAPILGYN-----LLDNSQKVLGDGATDTTNSDNLVNPVNNKGTETLPSTG 472
Db 452 ETKAPEGYNLOADPLPFRVLAKEAEK---KAATEVTITD-----VPKNAGFRPLTG 500
QY 473 GIGTTFYIIGALLVIGAGILVARRR 499
Db 501 ANGVIFLTAGALLVAGGAVVAYANKR 527

RESULT 10
Q9CHH3 PRELIMINARY; PRT; 614 AA.
AC Q9CHH3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Unknown protein.
GN YHGE OR LL0758.

```

OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX	NCBI_TaxID=1360;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=IL1403;
RX	MEDLINE=21235186; PubMed=11337471;
RA	Bolotin A., Wincker P., Mauger S., Jaillon O., Malarre K.,
RA	Weissenbach J., Ehrlich S.D., Sorokin A.;
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus
RT	lactis ssp. lactis IL1403."
RL	Genome Res. 11:731-753(2001).
DR	EMBL; AE006309; AK04856.1; ..
DR	PIR; F86719; F86719.
DR	InterPro; IPR008454; Cna_B.
DR	Pfam; PF05738; Cna_B; 2.
KW	Complete proteome.
SQ	SEQUENCE 614 AA; 61954 MW; B7DB40A29C53A5C6 CRC64;
Query Match	
Best Local Similarity 10.0%; Score 258.5; DB 16; Length 614;	
Matches 145; Conservative 57; Mismatches 184; Indels 207; Gaps 31;	
QY	51 VFDAEIDN-----ANVSDSNKDGASYLIPQGEAEYKASTDENSELFTTTTNGGRTYVT 103
Db	75 VNSGIANNTGGAGNTGKNDGSGQTTLSGSTPSTWANTFSAKYVGT-GVPTGVT 133
QY	104 K---KDTASANIATWAKSISANT-----TPVSTVTESNNDCGEVINYSQ----- 145
Db	134 DPSFTDTTSVAPVTTDASGLAFTGLTGQYLFHQVTTVNGITTVGDFIVQVSHEDSQAG 193
QY	146 ----YGYVYSTVNGAVIMTSVT-----PNATIE-----KNTD-- 178
Db	194 IVNVYPKLDMSSSAGLG-----TSATTNADNPNFNGQTPNOIANPNATGSDQTLNTDNN 248
QY	179 -----ATWGGGGKTVDOQTSYSGVDTVKYTIYKNAY-----NVHGTEKYV-QYV 222
Db	249 AGNEMLANGTWINGSDNQ-NTTAAAGNTVNNV---NTVFDSSQTNNGTGTGTGTGI 304
QY	223 IKDTMP-----SASVVDLNEGS-----YEVTTIDSGSNI--TTLTQ 257
Db	305 VTDQLPNNLVNSSTVTVSVIVNVINGSGTKVGLTPTDYIT-NDGNGKIWTLTAG 363
QY	258 SEKATGKNLLBENNFTIIPWAAATNPTGNTQNGANDFFYKGINITITVTVGLKSG 317
Db	364 QQHAASL--LGSADGALNIIIP-STVKSAGSATDSAT-----TTITNAY----- 405
QY	318 AKPGSADLPENTNIATINENTNDPGQKVTVRDQGITTIKKIDGSTKASLOGAIFVLKNA 377
Db	406 ----GADLSTTAV-----KSTLNVGGLMKTKDASTNAALAGATFTVRA 447
QY	378 TQO-----FLNFD--TNVWEGTEANATE-----YTTGAD 406
Db	448 DNKEDAQDFEANAAYFNNSAGGVTNLTSSKAAPVTDGTSNANTSATAPVTTTGD 507
QY	407 GIITITGLK-----EGT-----YVLVEKAPLGYNL-----LDNSQKVLGPGATDT 448
Db	508 GIATENGLNLVNDNTGSDNTNXYLVEVAAPGYQLPSVTTAANTLGAVTASTAPAAATD 567
QY	449 TNSDNLNPTVNNKGTLPSTGGTGTTFIYIIGAILVIG--AGILVARRR 499
Db	568 -----TITNNKFPALPTGGQ-----LAGIIAIVSGVIAFAIKR 604
RESULT 11	
Q81D71	PRELIMINARY; PRT; 553 AA.
ID	Q81D71
AC	Q81D71;
DT	01-JUN-2003 (TREMBLrel. 24, Created)
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE	Collagen adhesion protein.
GN	BC2508.
OS	Bacillus cereus (strain ATCC 14579 / DSM 31).
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX	NCBI_TaxID=226900;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	MEDLINE=22608415; PubMed=12721630;
RX	Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA	Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA	Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA	Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA	Overbeek R., Kyrpides N.;
RT	"Genome sequence of Bacillus cereus and comparative analysis with
RT	Bacillus anthracis."
RL	Nature 423:87-91(2003).
DR	EMBL; AE017006; AAP09468.1; ..
DR	InterPro; IPR001899; Gram_pos_anchor.
DR	Pfam; PF00746; Gram_pos_anchor; 1.
DR	TIGSFAM; TIGR01167; LPYTG anchor; 1.
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW	Collagen; Complete proteome.
SQ	SEQUENCE 553 AA; 60940 MW; 589F7B0E00187295 CRC64;
Query Match	
Best Local Similarity 23.8%; Score 255; DB 16; Length 553;	
Matches 144; Conservative 68; Mismatches 234; Indels 160; Gaps 26;	
QY	3 KKMIOSLVASLAFGMAVSPVTFPIAFAAETGTTVQDTQ----- 41
Db	2 KKFVLLVLFITFS-TWSSVLVADSPSKGTLTIHKYEQEKDQAGLSDGSANQEVK 60
QY	42 ----KGATYKAYKVPDAE-----IDNANVSDSNKDGASYLIPQ- 76
Db	61 DVKPLKGVTFEVRVASPEKISNDGKIVKEDVKPMGATPNQVTTDDNGQAVLKDLPLGR 120
QY	77 ---KEAEYKASDFN-SLFTT---TTNGRTYVTKOTASANEIATWAKSI---SANTT 125
Db	121 YEVREAGPPHVNLPNTTYTDIPLTNKEGVLNVDVHMYPKNEIKRGAVDLIKTVNEK 180
QY	126 PVSTVTES--NNDGTGEV-----INVSVQYGYVSVST-VNNGAVIMVTS-- 165
Db	181 ALAGAVFSLFKKDGEVKKELATDANGHIRVOGLEVEGYFOETKAPKGVVIDPTKREF 240
QY	166 VTPNATIEHNTDWTG-----DGGKTVDOK-----TYSVGDTVKYTIYKNVN 211
Db	241 VNSGTINEDGT-ITSGTVVKIEVKNNBEPTDKKINGKLEALPINPLTNVNYDIKTLIP 299
QY	212 YHGTEKVVQYVTKDTPSASVVDLNEGSVEVTITDG---SGNITTLTQSEKATG---KY 265
Db	300 -BDIKYKVVVTDLTDLNRLVI---QKPIVKI-DGAENVANVVEAIEGQKVATVKDF 354
QY	266 NLLBENNFTIIPWAAATNPTGNTQNGANDFFYKGINITITVTVTVGLKSGAKPGSADL 325
Db	355 TKLDGCKEFLQI-----KSGQKEGVPSG 378
QY	326 PENTNIATINPNTNDPGQK-----VTVRDQGITIKIDGSTKASLOGAIFVLKNATG 379
Db	379 SEILNTAKIHFTKNKDVIKESKPVWIPPTTGIIELTKIDSAKNKLGKAGFVLKDNNG 438
QY	380 QFLNFDNTNNWEGTEANATEVTTGADGIIITGLKEGYTVLVEKAPLGYNLLDNSQK- 438
Db	439 KIV-----VVAGKEVTG---VSDENGVIKWSNIPYGDYQIFETKAPT-YTKEDGTTKS 487
QY	439 -VILGDGADTTTNSDNLNPTVNNK-GTELPSTGGTGTTFIYIIGAILVIGAGIVLVA 496
Db	488 YQLKADPDKISENNQVTKLTIEKNKSGWILPVTGGIGTTLFTVTLGLTMLTAAPVFR 547
QY	497 RRRRLS 502
Db	548 KKFARN 553
RESULT 12	
Q48707	







```

QY 338 TSNDPQGVKVRDGOITIKKIDGSKASLOGAIFVLKNATQGFNLNNDNNVWEGTEAN 397
Db 369 PSN--PPRKEVHTGGRFVKKDSITETQLGAEFDLLASDGTAVKWTDA-LIKANTNKN 425
QY 398 --ATEYTTG-----ADGIITITGL-----KEG-----TYLVEKKAFLGYNLLDNSQ 437
Db 426 YIAGEAVTGPQIKLKSHTDGTETFEIKGLAYAVDANAEGTAVTYKLKETAPEGYVIPDKKEI 485
QY 438 KVLGDGA-----TDTTNSDNLVNPTVENNKGTLPSTGGIGTTFIYIIGAILV 487
Db 486 EFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTGGIGTALFVAIGAAMV 540

RESULT 16
Q820S9 PRELIMINARY; PRT; 554 AA.
AC Q8E0S9;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN SAG0845.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.B., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014220; AAM99537.1; -.
DR TIGR; SAG0645; -.
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
KW Complete proteome.
SQ SEQUENCE 554 AA; 60168 MW; 64498D4B49729AD1 CRC64;

Query Match 9.5%; Score 244.5; DB 16; Length 554;
Best Local Similarity 25.4%; Pred. No. 0.0011;
Matches 151; Conservative 61; Mismatches 218; Indels 165; Gaps 31;

QY 1 MKKKMIOSLLVASLAFGNMSPVTPPIAFAETGTTV-----QDTQKGATYKAYKV--- 51
Db 3 LSKKLLFSAAVLTVAGSTVEPVAQFA---TGMISIVRAAEVSQERPAKTVNIYKLQAD 58
QY 52 -FDAIDNANVSDNKGQ--ASYLIPQKEAEYKASTDPNSLFTTTTNGGRTYVTKKDTA 108
Db 59 SYKSEI-TSNGGIENKQGEVINSYAKLGDNVKGGQGVQF-----KRYKVKTD-I 105
QY 109 SANEIATWAKSIGISANTTTPVSTVTE-----SNNDGTEV---INVSQGYGVVVSSTVNN 157
Db 106 SVDELK-LTTVEADAKVGTILLEGVSLPKQTNAGVLVDALDSKNVRVLYVEDLKNs 164
QY 158 GAVIMVTSVTP-----NAT-----IHERN-----TDATWGDGGGKTVQDKTYSVD 198
Db 165 PSNITKAYAVPFVLELPVANSTGTGFLSEINIYPRNVVTDPEKTKDVKVLGQ-----D 218
QY 199 TVKYTIITYKNVNHGTEKVQYVIKDTWPSASVVVDLNEGSYE-VTITDGSNITLTITQ 257

```

```

Db 219 DAGYTI-----GEE--FKWFLKSTIPA-----NLGDYKFEITDKPADGLTY--- 258
QY 258 SEKATGKYNL-----LENNNFTIIPWAAATNPTGTQNGANDDF-----FYKGI 303
Db 259 --KVGKIKIGSKTILNDEHYTI-----DEPTVDNQNTLKITFKPKPKFAIELLKG 309
QY 304 NITIVTYTVGLKSGAKPGSADLPENTNIATY-----NPN 337
Db 310 -TLVKNQDALDKATANTDDAAFLPVPVASTINERKAVLGKAIENTFELQYDHTDPKADNPK 368
QY 338 TSNDPQGVKVRDGOITIKKIDGSKASLOGAIFVLKNATQGFNLNNDNNVWEGTEAN 397
Db 369 PSN--PPRKEVHTGGRFVKKDSITETQLGAEFDLLASDGTAVKWTDA-LIKANTNKN 425
QY 398 --ATEYTTG-----ADGIITITGL-----KEG-----TYLVEKKAFLGYNLLDNSQ 437
Db 426 YIAGEAVTGPQIKLKSHTDGTETFEIKGLAYAVDANAEGTAVTYKLKETAPEGYVIPDKKEI 485
QY 438 KVLGDGA-----TDTTNSDNLVNPTVENNKGTLPSTGGIGTTFIYIIGAILV 487
Db 486 EFTVSQTSYNTKPTDITVDSADATPDTIKNNKRPSIPNTGGIGTALFVAIGAAMV 540

RESULT 17
Q9X4D1 PRELIMINARY; PRT; 535 AA.
AC Q9X4D1;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Type-1 fimbrial major subunit precursor.
GN FIMP.
OS Actinomyces viscosus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Actinomycetaceae; Actinomycetaceae; Actinomycetes.
OX NCBI_TaxID=1656;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19246;
RX MEDLINE=99242779; PubMed=10225854;
RA Li T., Johansson I., Hay D.I., Stromberg N.;
RA "Strains of Actinomyces naeslundii and Actinomyces viscosus exhibit
RA structurally variant fimbrial subunit proteins and bind to different
RA peptide motifs in salivary proteins.";
RL Infect. Immun. 67:2053-2059(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC19246;
RA Li T.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF106034; AAD28826.2; -.
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Signal.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 535 TYPE-1 FIMBRIAL MAJOR SUBUNIT.
SQ SEQUENCE 535 AA; 56378 MW; E64F5B13A00F1E7 CRC64;

Query Match 9.4%; Score 243; DB 2; Length 535;
Best Local Similarity 23.6%; Pred. No. 0.0013;
Matches 133; Conservative 71; Mismatches 232; Indels 128; Gaps 25;

QY 13 SLAFGNMSPVTPPIAFAETGTTVQDTQKGATYKAYKVFDAIDNANVSDNKGASYL 72
Db 17 TLAVGALIAIP-TGAAAPADPNGSTIDPETH-ATLTVHKCEQTDNNGVKEGTGNEPTT-- 72
QY 73 IPOKKEAEYKASTDFN-----SLFTTTTNGGRTYVTKKDTASANEIAT-WAKSISANTTP 126
Db 73 -----ECNPVSGVEFTITRLNVDLTTDEGWNKTLAALQGNVAGQAUKTQTSITTTGADG 127

```

```
QY 127 VSTVTESNDGTEVINVSQGYVYVSTVNGAVIMVTS--VTPNATHEKNTDATWG-- 182
DB 128 LASFTDAQTD-----VGAYLVSETRDPKVI PAEDFVVTLPMNTPEKTEWYNVH 178
QY 183 -----DGGGKTVQDKTYS-VGDTVKYITIKY-NAVNVHGTEKVVQYVVKDTM----- 227
DB 179 VPKVTVSVDVQVSDKQVSGSGNDITYTITTSIPKVDYAGGARIKRYEVVDOLDKRIKK 238
QY 228 -----PSASYVDLNEGSYEVTITDGSNIITTLQSGSEKATGKYNLLSENNFTITIPWAAT 283
DB 239 DQLTPVVKIIGAAGNPEITLVDGTDYTVITADGANHNWATIQLTTEGRKAA-----EAR 294
QY 284 NTPGTONTGANDDPFKGINITVITYT-----GVLSGAKPGSGADLPENNTIAT 333
DB 295 ANGTGETK-----VQVTLTAKFSDVDLEGLTSLNTAGLIPSDSP-NPDMWP 339
QY 334 INPNTSNDPPGQKVT---VRDGOITIKKI---DGSFKASLQGAIFVL-----KNATGOFLN 383
DB 340 SNPGT--DVPGTPTTISKYKVNLTGTGTDNLADKTKYNGAEFQVYECTKTATGATLK 397
QY 384 FND---TNNVWGTTEANATEYTTGADGIIITIGLK-----EGTYILVEKKA 426
DB 398 DADASTADKVDPLTIGGKQTFTEGQGMVSIATLRANDVYNGAEKQLTDDDDYCLVETKA 457
QY 427 PLGYNL-----LDSQKVLGDGATDTTNSDNLVNPTVNNKGTSLPSTGGIG 475
DB 458 PEGYTLQADPIPFVRVLAADAERKI-----PTEVTVTD-----VPKAGFRLPLTGANG 505
QY 476 TTFIVFIIGAILVIGAGIVLVARRR 499
DB 506 VIPLTIAGALLVAGGAVAYANKR 529

RESULT 18
QBE4C4
ID QBE4C4 PRELIMINARY; PRT; 674 AA.
AC QBE4C4;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN GBS1477.
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513 (2002).
DR EMBL; AL766851; CAD47136.1; -.
DR SAGAList; gbs1477; -.
DR InterPro; IPR008454; Cna_B.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna_B; 2.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 674 AA; 73218 MW; 211133B59FC4727D CRC64;

Query Match
Best Local Similarity 21.4%; Pred. No. 0.0049;
Matches 149; Conservative 82; Mismatches 217; Indels 249; Gaps 32;
QY 9 LLVASLAFGMVSPVTPIAFAAETGTTVQDTQKGATYKAYKVFDAEIDNANVSDSNKDG 68
```

```
DB 16 LILTSL---FSVAP-----AFADDVTTDTV-----TLHKIVMPQAAFNFTEGTKGKD 61
QY 69 ASYLIPOQKE-----AEYKASTDFNSLFTTTTNGGRYVTKKDTASANEIATWAKSISANT 124
DB 62 SDYV---GQINDLSKSGSTDAKEI-----KGAFVFKNETGT-----KFTENG 104
QY 125 TPVSTVTSNNDGTVIN-----VSQGYVYV-----SSTVNGAVIMVTSV 166
DB 105 KEVDTLEAKDAEGGAVLSGLTKDTGFAFNATKLKGTQIVELKEKSNYDNNGSILADSKA 164
QY 167 TP-----NATHEKNT-----DATWGGG-----GKTV-DQ 191
DB 165 VPKVITLPLVNNQGVVVKOAHYIPKNTETKPKQVDKFNADKOLDYDNRKDKGVVSATVGGK 224
QY 192 KTVSYG-----DTVKYITIKYKNAV-----NYH--GTEKYVQV 222
DB 225 KEYIVGTLKGSYDKLVWTDSTMTKGLTFNNNVKVLTDGKDFPVLYNKLVTDDQGFRLA 284
QY 223 IKDTPSASVDLNEGSYEVTIT---DGSNIITTLQSGSEKATGKYNLLSE----- 270
DB 285 LNATGLAAVAAAADKDVETIKITYSATVNGSTTVEVPETNDVKLDYGNPTSESEPEQGT 344
QY 271 ---NNFTITIPKAATNPT-----GNTONGAND-----DPFYKG 302
DB 345 PANQEIWKIDWAVDGTITDNNVAVKAITLQEKQTDGTWVNVASHEATKPSRFEHTFTG 404
QY 303 INTIITVTVTVGLK-SGAKPGSADLPENTNIATINPNTSNDP-----CQKVT 348
DB 405 LDN-TKTYRVVERVSGTYPEYVSF--KNGVVTIKNNKNSNDPTPINPSEPKVVTYGRKF- 460
QY 349 VRDGOITIKKIDGST-----KASLQGAIFVLKNATGOFLN----- 383
DB 461 VKTNQANTERLAGATFLVKKEGKYLARKAGATAAKAAVKTAKLALDEAVKAYNDLTKE 520
QY 384 -----FND-----TNNVWGTGA---NATVTTGADGIIITIGLK 415
DB 521 KOEGEGKTALATVDOKAYNDAPFKVANSYEWVADKADNVVKLIISNAGQGFETGLD 580
QY 416 EGTYYLVEKKAPLGYNLDNSOKVILG-----GATDTTNSDNLVNPTVE-- 461
DB 581 KGTYSLEETQAPAGYATLS-----GDVNFVETATSYSGKATTDIADKSGVKDAQOV 633
QY 462 NNKGTELPSTGGIGTITFIYIIGAILVIGAGIVLVARR 498
DB 634 QNKVVTIPQTGGIGTITFIYIIGAILVIGAGIVLVARR 498

RESULT 19
QBYAG7
ID QBYAG7 PRELIMINARY; PRT; 793 AA.
AC QBYAG7;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DE Putative peptidoglycan bound protein (LPXTG motif).
GN LMO0159.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek B., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
```



```
GN LMO2085.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591982; CAD00163.1; -.
DR PIR; AE1335; AE1335.
DR ListLiSt; LMO2085; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 3.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 562 AA; 60457 MW; F7485EB9B47DDPED CRC64;

Query Match      8.4%; Score 217; DB 16; Length 562;
Best Local Similarity 23.2%; Pred. No. 0.02;
Matches 131; Conservative 55; Mismatches 197; Indels 182; Gaps 27;

QY 32 TGTITVQDTQKATYKAYKVFDAEDNANVSOSKDGASYL-----IPQKEAYKAST 85
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 80 TGNIEVWQNGYNPAP-----SGNFIELNSDGIGPVYQDIRTIP-GSNLTWKS- 130
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 DPNLSFTTTTNGRTYVTKDITA-----SANEIA-----TWAKSISANTT----- 125
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 131 -----HGRGTGV-----DTADLLIGSPESQTEVSRVSNGETWG-SFEGNYTVPAQOT 176
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 -----PVSTVTESNNDCGTEVINVSQYGYVVSSTVN-NGAVIMVTSVTPNATIHKNT 177
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 177 ITRLTFNPISTANGSLTSGNFLDDVQLY-----INVNGA-----KIG 213
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 178 DATWGDGGGKTVQKTYSGVDTVKYTI-----TYKNVNYHGTQYQVVIKDTWPSASV 232
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 214 DVVWYDFNGDGIQDSEBPAGVKVYDLTLTKDGTFKESAT---TNNIGSYLFTDVLV- 266
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 233 VDLNNGSSEV---TITDGSNITTLTGSEKA-----TGKYNLLEEN---NNFTITIPW 280
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 267 -----GDYQVKFSPNNDPIFSKANQGNKLSNKPDKTGFIASVNPVLKSNFQI- 317
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 281 AATNTPGTNTQNG-----ANDDPFYKGINITITVYTG-V-LKSGAKPG 321
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 318 -----DAGITNGKVEIQKLSGDKALSGAVAIKNSQSEVAKITNQNGTGTAEGLPFG 372
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 322 SADLPENTNIATINPTS-----NDDPQKQVTRD-----GQITIKKIDGSTKASLQG 369
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 373 NYTATEVTAPLGYQKNTTPKFTITYGDTNPKVLTQFNAEKTSITIPKQDEANKKGLAN 432
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 AIFVLKNATGQELNFNDNNVWEGTEANATEYTTGADGIITITGLKEGTYILVEKKAPLG 429
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 433 AVFDKSIDGTLK-----KVTTSKGYALAEINLQPGYVITEVAPP 476
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 YNLLNSQKVLGDGATDTN---SDNLLVNP-----TV-----ENNKGTLPSTG 472
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 477 YEKSAEIRVTIPFPNQKNTINTFSDNKLIMVLPKPTPKGTVKVSGETTKITALPQTG 536
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 473 GIGTTFIVIIIGAILVIGAGIVLVAR 497
DB 537 DSSSSSTIFTGLLIIVVAGSLFVYRR 561

RESULT 22
Q929J3 PRELIMINARY; PRT; 1806 AA.
AC Q929J3;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative peptidoglycan bound protein (LPXTG motif).
GN LIN282.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596171; CAC97510.1; -.
DR PIR; AF1717; AF1717.
DR ListLiSt; LIN282; -.
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR008456; Collagen bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 8.
DR Pfam; PF05737; Collagen_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1806 AA; 195670 MW; ABC808A65D84972F CRC64;

Query Match      8.4%; Score 216; DB 16; Length 1806;
Best Local Similarity 23.8%; Pred. No. 0.083;
Matches 114; Conservative 50; Mismatches 168; Indels 148; Gaps 19;

QY 43 GATYKAYKVFDAEI-----DNANVSDSKDGASVLIPOGKEAYKASTD--- 86
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 721 GKGAYNSTGNIDWIVSANAMAKSYDNLIPDDTPTGLTYVEGSLQVRNVESSTEMS 780
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 87 ----FNSLFTTTNGRTYVTKKOTASANEIATWAKSISANTTPTVSTVTESNNDCGTEVI- 141
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 781 LQIPLNSTGLAKAGDKYKPKVDI-TGNKI-----HLEFALNDTRVFI 824
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 -----NVSQGYGYVTSVTVNGAVIMVTSVTPNATIHKNTDATWGDGGGKTVQKTYVS 195
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 825 KYSTKPDENWYFYKYVTN-----IAEYTDNGT-----DKHTYS 857
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 196 -----VGDYVKYTIYKNAVNYHGT-----EKYQYVVIKDTMPSASVVDLNE 237
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 858 YQAYASKVFNAMTKASIDSTYNNKVNMTVTLLNISKORPINNPTITDTM-----KN 909
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 238 GSYEYTIITDGSNITTLTGSEKATGKYNLLEENNNFTITIPWAAATNTPTGNTONGAND 297
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 910 GTTGQAVQVKSFRVINEITGDDIDSKYYDIITFDKDFIQFK-----951
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 PFYKGINITVYTG-V-LKSG-----AKPGSADLPENTNIATINPTSNDPQKQV 348
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 952 -DYKATDPIKVTYSTVSLMSGLISNTATTATSPDYGSLPMSYKSRRTNISPAFTIGSGGT 1010
QY 349 VRDQITIKKID--GSTKASLQGAIFVLKNAATGQFLNFDNTNNEVGTETANATEYTTGAD 406
Db 1011 ATTGSLEITVKDKDNFK-KLTAKAFQLYTPEGD-----KAGQEA-----TTDSE 1054
QY 407 GIITITGLKGTYYLVEKKAPLGYNLLDNSQKVLGDGATDTTNSDNLLVNPVTENNKGTT 466
Db 1055 GKIWMDSLQSGKYKLVEAAPTGTISDEYK-----DGKEITITAD-VATSVTIENTEQT 1108

RESULT 23
Q89PB9 PRELIMINARY; PRT; 3441 AA.
AC Q89PB9;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE BL13563 protein.
GN BL13563.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Iidesawa K., Iriuchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005948; BAC4828.1;
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004601; P:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
DR InterPro; IPR001343; Hemiyen Ca bind.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00353; hemolysincabin; 7.
DR PRINTS; PR00313; CABNDNGRPT.
DR PROSITE; PS00435; PEROXIDASE_1; 2.
KW Complete proteome.
SQ SEQUENCE 3441 AA; 342613 MW; D7175RCB122D0B82 CRC64;

Query Match 8.4%; Score 215.5; DB 16; Length 3441;
Best Local Similarity 23.6%; Pred. No. 0.18;
Matches 115; Conservative 60; Mismatches 165; Indels 147; Gaps 25;

QY 27 AFAAETGTTIVDTQKATYKAYKVPDAEIDNANVSDSKDGASYLIPOGKEAEYKASTD 86
Db 2200 SFNAATGTLTSGSSVANYQT--ALDS-VTYLNTSD-NPSGLARTV-----2242
QY 87 FNSLFTTTNGRYYTKDTSANAEIATWAKSISANTTPVS-----TVTESNND 136
Db 2243 -----TIITNDG-----TANSVAV---TDTINVPVNDAPGVNDTGSATDAGT 2284
QY 137 GTEVINVSQGYGYVYST-VNNGAVIMVTSVTPNATIHENKNTATWGDGGGKTVDOKTYS 195
Db 2285 NGTAGSNATGNVLTNDTDVNTNLSLVSAIRTGAVEGSGTAGTLG-----2331
QY 196 VGDVVKYTYTKAVNVHGTGKV-----YQYVLKDTMPSASVVDLNEGSVEYVITDGSN 250
Db 2332 -----SILVAGHTLTINANGSYTYVND--DSAVQALNSGQ---TITD-SFN 2374
QY 251 ITTLTGSEKATKYNLLNENNFITTI PWAATNPTGTONGANDDPFYKGINITITVY 310
Db 2375 YAVKDPGN-----LTDTAULTVTI-----NGAND-----APNVTV---2404
QY 311 TGVLKSGAKPGSADLPENTNIATNP-----NTSNDPDGQ-----KVTVRDQGQITIKKIDG 361
```

```
Db 2405 -----PGTQEVVQNTNV-TFNGAKLISISDVVDVGAGTETVTLVAHGTLTSGTTG 2454
QY 362 SYKASLQGAIFVLKNAATGQFLNFDNTNNEVGTETANATEYTTGADGIIIT-----GLKE 416
Db 2455 LSFETGDTGTTDATTMTSGTAANNALN---GLLYNPDTDTFVGAD-TLTITTTDQGLSD 2510
QY 417 GTYLVKKAPLGYNLLDNSQKVLGDGATDTTNSDNLLVNPVTENNKGTELPLSTGGIGT 476
Db 2511 SDTIVINQSPNGITLTSSDVI FYASGNTVNATNLIN-----GTD-SITGGTGT 2562
QY 477 TIFYIIG 483
Db 2563 DTLIVTG 2569

RESULT 24
Q92FA4 PRELIMINARY; PRT; 681 AA.
AC Q92FA4;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Surface anchored protein (LPXTG motif).
GN LINO202.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Duesurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkak G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Norddick G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species."
RL EMBL; AL596163; CAC95435.1;
DR PIR; AC1458; AC1458.
DR Listlist; LINO202;
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR008456; Collagen bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 3.
DR Pfam; PF05737; Collagen_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
KW Complete proteome.
SQ SEQUENCE 681 AA; 73391 MW; C9B713CFE1547FD0 CRC64;

Query Match 8.3%; Score 214.5; DB 16; Length 681;
Best Local Similarity 22.0%; Pred. No. 0.033;
Matches 110; Conservative 56; Mismatches 194; Indels 141; Gaps 16;

QY 17 GMAVFPVTPIAFAAETGTTIVDTQKATY-KAYKVFDAE-----IDNANVSDSNK 66
Db 12 GLAFLSVLIATVQTIVKATSYGSEFLNTVELLDKDGTPKTDGYYDNVNH-----67
QY 67 DGASYLIPOGKEAEYKASTD-----NSLFTTTNGRYYTKDTSANAEIATW 116
Db 68 --YTWISIPNSTNVKAGDSMDPALSQLALATDLAFNVKDSKQTVGTATVKRATNQV---122
QY 117 AKSISANTTPVSTVESNND-----GTEVINVSQGYGYVYSTVN-NG 158
Db 123 -----TVVPSDYVEKSHDKGELDFWTFNQKITGNEKINLE---FPIENSTINVDV 172
```

```
QY 159 AVIMTSVTPNATIHKNKNTATWGGG-----KTVDQKTSVSGDVTVKYTI 204
DB 173 NVGEKTPVSPTEIFPKYG-----WVDANNPSLIHWVRVNYAQNIPNAVFDTIIGAKQTL 228
QY 205 TYKNAVNVHGTGKTVQYVVIKDTMPSASVVDLNEGSEYVITDGSNITTLTGGSEKATGK 264
DB 229 NFDISKAFHGTVSTDRVTAGTPIENTFESKTSDFSVTL-----GNLDSVQIS----- 278
QY 265 YNLLENNNFTITIPMAATNTPTGNTONGANDFFYKGINITITVYTGVLKSGAKPGSGAD 324
DB 279 -----YTTATDGGKSTQYDNTAKLSGEDFTVKOTSTWTPASGG---GGGADGTT- 325
QY 325 LPENTNIATINPNTSNDPQKQVYVVDGQIITIKKIDGSTKASLQGAIFVLKNATGQFLNF 384
DB 326 -----GSVLTIKQDAKTATKATLEGAEPKFLVDANGTILQOE 358
QY 385 NDTNNVWTEANATEYTTGADGIITGLKEGTVYLVKPKAPLGVNLLDNSQKVLGDDG 444
DB 359 N-----ITTDASQLNVSNLKPATYQLTIETKAPTGYKLDSTPVEFTIGEN 403
QY 445 ATD-TTNSDNLVNPVTENK 464
DB 404 NKEITVTKENTLDTGAVELTK 424

RESULT 25
Q8CY18
ID Q8CY18 PRELIMINARY; PRT; 2551 AA.
AC Q8CY18;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SPRI403.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
McAhen S.M., McHenry M., McLeaster K., Mundy C.W., Nicase T.I.,
Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008509; AAL00207.1; -.
DR PIR; B98047; B98047.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001298; Filamin.
DR InterPro; IPR005877; Gpos_Y5IRK.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF01391; Collagen; 10.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF04650; Y5IRK signal; 1.
DR ProDom; PD000007; Clg helix; 1.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR TIGRFAMs; TIGR01168; Y5IRK signal; 1.
DR PROSITE; PS50194; FILAMIN_REPEAT; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 2551 AA; 265126 MW; B3A55D0AE30DA146 CRC64;
```

Query Match 8.2%; Score 212; DB 16; Length 2551;  
Best Local Similarity 23.3%; Pred. No. 0.19;

```
Matches 121; Conservative 57; Mismatches 194; Indels 148; Gaps 24;
QY 34 TITVQDTQKATYKAYKVFDAIDN-----ANVSDSNKDGASYLI 73
DB 1891 TFTVKGKNGKDGKAPKIKVEDITSPIRIRDTAAATPRNGIRVTYVDDVNDNGV-- 1948
QY 74 PQKRAEYKASTDFNSL-----FTTTNGGRTYVTKDTASANEATWAKSISANT 124
DB 1949 DEGVDKLVNKSIDYNGIDGRGSAFTITTKONGDGTHTITVQNPDSSESTTVVKGKQK 2008
QY 125 TPVSTVTSNNNDGTENVINVSQYGVYVSVTVNNGAVIMVTSVTPNATIHKN-----TDA 179
DB 2009 TANITTE-NPDGSHITVTPNPDGSKETVVGKNGK-----DGKTPKVEYTDNDGTHTVKV 2063
QY 180 TWGDCG-----GKTVDQKTSVSGDVTVKYTYTITKNAVNYHGTGKYYVYIKD--- 225
DB 2064 TDGDNVTNNAIKGKDGKAATATTENPDG-SHTVITNP---DGTKN---EFVVGKGRD 2117
QY 226 ----TSPASVVDLNEGSEYVITDGSNITTLTGGSEKATCKYNLLBENNFTITIPWA 281
DB 2118 GVDGRTPTASVRDNGDGSHTIVITNPEG-VTTEITVRDGSKPKVTITDB----- 2165
QY 282 ATNTPGTONTGANDDDFFYKGINITITVYTGVLKSGAKPGSADLPENTNIATINPNTSD 341
DB 2166 -----QNGTHKISVLNGDGTETET---IIGKSPVATVRDNDGTYTIRVENGNG 2213
QY 342 DPGQKVTVRDGQ-----ITIKKIDGSTKASLQGAIFVLKNATGQFLNFND 386
DB 2214 TVSE-TTVRDGKSPAKVVDNGDGTHTITVNSDGIITTT-----TVRDGKPKLEVID 2266
QY 387 TNNVWGTTEANATEYTTGADGIITGLKEGTVYLVKPKAPLGVNLLDNSQ----- 437
DB 2267 NN-----DGSHTIKVTKGADGKGTITTFIDG-----KSPKA-NIVDNGDGTHTLTIVD 2312
QY 438 -----KVILGDGATDTTNSDNLVNPVT---ENKGTTEL 468
DB 2313 SDGREYKSIKDG-KDGKDS-----VSPITVTVKNNNDGTHV 2347

RESULT 26
Q836L9
ID Q836L9 PRELIMINARY; PRT; 476 AA.
AC Q836L9;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Cell wall surface anchor family protein.
GN EF1092.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=13951;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=V583 / ATCC 700802;
RX MEDLINE=22550857; PubMed=12663927;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
Daghighy S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nealeon W.,
Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
Utterback T., Radune D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AE016950; AAO80892.1; -.
DR TIGR; EF1092; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW Complete proteome.
SQ SEQUENCE 476 AA; 53355 MW; 1D957EDD3009D99D CRC64;
```

Query Match 8.1%; Score 208.5; DB 16; Length 476;

Best Local Similarity 26.5%; Pred. No. 0.041;		Matches 85; Conservative 49; Mismatches 124; Indels 63; Gaps 14;	
QY	187 KTV--DQKTSVGDVTKYITTKYKNAVNYHGTGVQVYVWIKDTWPSASVVDLNEGSVEVTI 244		
Db	192 KTVLDKQGFNGQEHINQYQITQIPANILG---YQ-----EFL 227		
QY	245 TDGSGNITLTQGS--EKATGK-----YNLLENNFTITIPWAAATNTPTGNTQNGANDD 297		
Db	228 SDRKATTLTLLPESIEVKVAGKVTVTGTYLTQKHGFTL-----DFSIKDLQNFAN-- 278		
QY	298 FFYKGINITVTVTVGLKSKAGPGSADLPENTNINATINPNTSNDPQGVKTVRQGITIK 357		
Db	279 -----QMTVSYQMLEKTAEPDTAINNEGQLV-----TDKHTLTKRATVTRGKSFV 326		
QY	358 KIDG-STKASLOQAIFVLKNATQGFNFNTNNVEWGT-ANATEVTTGADGIIITITGLK 415		
Db	327 KVDSENAKITLPFAVIVKQAGEYN-ETANGYRWQKEKALAKFTSQAGEFSVGLK 385		
QY	416 EGPYLYVEKKAPGLYNLLDNSQKVLGDGATDNTSNDLLNPNVTNNKGT-----LPSTG 472		
Db	386 DGQYFLEEISAPKGYLLNOTEIPFTVGKNSYATNGORTAPLH--VINKVKESGFLPKTN 443		
QY	473 GIGTTFYIIGAILVIGAGIV 493		
Db	444 E-ERSIWLTIAGLLIIGWVI 463		
RESULT 27			
Q88UJ0	PRELIMINARY; PRT; 917 AA.		
AC	Q88UJ0		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)		
DE	Cell surface protein precursor, GY family.		
GN	LP_2486.		
OS	Lactobacillus plantarum.		
OC	Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;		
OC	Lactobacillus.		
OX	NCBI_TaxID=1590;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=NCIMB 8826 / WCFS1;		
RX	MEDLINE=22480296; PubMed12566566;		
RA	Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,		
RA	Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,		
RA	Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,		
RA	Hoffer S.M., Nierop Groot M.N., Kerhoven R., De Vries M., Ursing B.,		
RA	De Vos W.M., Siezen R.J.;		
RT	"Complete genome sequence of Lactobacillus plantarum WCFS1.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).		
DR	EMBL; AL935259; CAD64786.1; -		
DR	InterPro: IPR001899; Gram_pos_anchor.		
DR	Pfam; PF00746; Gram_pos_anchor; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 917 AA; 91739 MW; 33449A470900648B CRC64;		
Query Match		8.1%; Score 208; DB 16; Length 917;	
Best Local Similarity 23.0%; Pred. No. 0.09;			
Matches 133; Conservative 56; Mismatches 207; Indels 182; Gaps 29;			
QY	27 AFAAETGTTT-----VDQTKGATYK-----AYKVF-----DAEIDNA 59		
Db	187 AGSVDAAGTILTSATQSGGSLQDQAQYIQENVDTNKKVYTAGTYAVARLTKSQALDWA 246		
QY	60 NVS-----DSNKDG-----ASYLLPOGKEAEYKASTDFNSLFTTTTNGRGTYVTKDT 107		
Db	247 NASGQVSIAPNGNSNGTWAVEAYTESKEYSAAAGASTATVDITKLTDADSYVTVLT 306		
QY	108 ASANEIATWAKSTS-----ANTTPVSTV--TESNN-DGTEVINVSQYGYVVSQVTVNNGAVI 161		
Db	307 FXANDATTGSRAAVLEFTGTTISVKNLSTNTNNTDANQOIEAWSYATQVMDTSVAGTVV 366		
Best Local Similarity 26.5%; Pred. No. 0.041;		Matches 85; Conservative 49; Mismatches 124; Indels 63; Gaps 14;	
QY	162 MVTSTPNATIIHEKNTDATWGGGKTVDPQKTSVGDVTKYITTKYKNAVNYHGTGVQY 221		
Db	367 -----VHVVD-----ENGKIAADTTVQGDVNDTVTTPATPSNY----- 401		
QY	222 VIKDMPASVVDLNEGSVEVTITDGSNITLTTOGSEKATGKYNLLENNFTITIPWA 281		
Db	402 -LDTTKKSALT-----GIVAADTTDSGNTV--AAGTE-LTLVYSQNTASNLTVNYVDA 453		
QY	282 ATNT--PTGNTQNGAN-----DGF-----FYKGINITVTVT 311		
Db	454 DGNITLPSKTYTEGADGTAEEVGGAYSVNAASIDGYTLTGDATAQTGTTFVSGGNTVTFVT 513		
QY	312 GVLKSGAKPGSADLPENTNINATIN-----PNTSNDPQGVKTVRQGITIKKID 360		
Db	514 ---KDA-----PVEQSTVTVNYVDADGNTIKAATQTOTLDNGSYTYVE-----TPTID 558		
QY	361 GSTKASLOQAIFVLKNATQGFNFNTNNVEWGTBANATEYTT-----GADG-----IIT 410		
Db	559 GYTKSADAAL-----TGT-VDGNKTTLTITTKNATPVEQSTVTVNYVDADGNTIKAAT 611		
QY	411 ITGLKEGTYYLVEKKAPLGY-----NLLDNSQKVL-----GDGATDTTNSDNLVN 457		
Db	612 TOTLDNGSYTYVETPTIDGYTKSADAALTGVDGNKTTLTITTKDSTPTTPVENKANLTIN 671		
QY	458 -----PTVENNKGTLPSTGGIGTTFIYIIGAILVIG 489		
Db	672 YVDADGNTIKASSVTE-----YIVGQAYTVG 697		
RESULT 28			
Q8FMP6	PRELIMINARY; PRT; 520 AA.		
AC	Q8FMP6		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DE	Putative type 1 fimbrial protein.		
GN	CE2457.		
OS	Corynebacterium efficiens.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacterineae; Corynebacteriaceae; Corynebacterium.		
OX	NCBI_TaxID=152794;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;		
RA	Kawarayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,		
RA	Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,		
RA	Usuda Y., Sugimoto S.;		
RT	"The entire genomic sequence of Corynebacterium efficiens YS-314.";		
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AP005222; BAC19267.1; -		
DR	InterPro: IPR001899; Gram_pos_anchor.		
DR	Pfam; PF00746; Gram_pos_anchor; 1.		
DR	TIGRFAMS; TIGR01167; LPXTG_anchor; 1.		
DR	TIGRFAMS; TIGR01409; TAT_signal_seq; 1.		
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 520 AA; 54823 MW; 4ABB44EF0762B05B CRC64;		
Query Match		8.0%; Score 206; DB 16; Length 520;	
Best Local Similarity 22.9%; Pred. No. 0.058;			
Matches 131; Conservative 68; Mismatches 200; Indels 172; Gaps 28;			
QY	14 IAFGNVSPVPIAFAAETGTTVDQTKGATYKAYKVFDAEIDNANVSDSNKDGASYLI 73		
Db	32 LAKAGALLAAALAFSPASPSVAQTNDVPETA-----NVVITKLEQPTVAGSVASGQSQT 86		
QY	74 PQG-----KEAEYKASTDFNSLFTTTTNG-----GRTYVTKKDTASANEIATWAKS-ISANT 124		
Db	87 PEGSKGIEDVEFTVR--VPVYKAGENGDEHPARTSAMQOATADID--LDMAKSRVAELA 142		



```
QY 125 TPVS-TVTESNDGTEVINVSQGYVYSSVYNNGAVI-----MVTSTVPNAT----- 171
DB 143 TPEVALNPTENGVNTNNSAAAAGLYLIRETSTPAGVIPADQFLLAVPMTNPEGTGWLT 202
QY 172 --IHEKNTDATMGDGGKTVQD-KTVSVGDTVKYTIYKNAVNYHGTEKVYQYVVKDTP 228
DB 203 IYVYPKGSTFS-----ASKSVDNAASLKTDGDPVTWIE-----AGIPLIRSHTSQDPA 251
QY 229 SA--SVVDLNEGSYEVTIDSGNITTLTQSEKATGKYNLLENNNFTITIPWAATNP 286
DB 252 PAEFKIVD--FENTADLELVGGANGVTVPVTEFTVADY-----TVDV----- 293
QY 287 TQNTONGANDDFYKGINITITVY-----TGVLKSGAKPGSADLPE 327
DB 294 -----NPVVD-----GLTTTITFNEGGKLVGKDDVTVELKTTVLKAG-----E 334
QY 328 NTNIAINPNTSNDPPGQ-----KVTVRDGOITIKKIDGSTKASIQAGIFVLKNGATGF 381
DB 335 INTTATI---TATDDTRATVTVDIKQTVKYGVTLTK-----KNQNGQL 375
QY 382 LNFNDNNVWTEANA-----TEYTTGADGI 408
DB 376 A--KDAVFRVYATEAHAKAGNDHLITATNKTGWTGGKVLGDGFRFSDFADGADQV 433
QY 409 IITGLKECTYVLVEKAPLGNLDNSOKVILGDGATDTTNSDNLNPTVNNKGTTEL 468
DB 434 KGADGSLYQTYLWLVEMTAPAGQQL--AEPVFP-----TVTEVEDTLEVTNTA--NTNAFVL 486
QY 469 PSTGGIGTTFIYIGAILVIGAGIVLVARRR 499
DB 487 PUTGGTGTAMLTILG--IGILAIVLFVARSR 515

RESULT 29
Q929J4
ID Q929J4 PRELIMINARY; PRT; 1622 AA.
AC Q929J4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Probable cell surface protein (LPXTG motif).
GN LIN2281.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fslhi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerel U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitoullam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596171; CAC97509.1; --
DR PIR; AE1717; AE1717.
DR ListList; LIN02281; --
DR InterPro; IPR008454; Cna B.
DR InterPro; IPR008456; Collagen bind.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF05738; Cna B; 6.
DR Pfam; PF05737; Collagen_bind; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
```

```
KW Complete proteome.
SQ SEQUENCE 1622 AA; 178366 MW; 7A12AS775BA6FF37 CRC64;

Query Match      8.0%; Score 205; DB 16; Length 1622;
Best Local Similarity 22.4%; Pred. No. 0.23;
Matches 134; Conservative 65; Mismatches 207; Indels 192; Gaps 27;

QY 18 MAVSPVTPIAFA---AETGTTIV-----QDTQK---GATYKAYKVFDIAEDIANVSDSNK 66
DB 991 VAVSPFTVGTGSGIATIGTIKITKVDDEDTTKLEGAKFQLY-TLDGEKSGQEIT-TNS 1048
QY 67 DGASVLT--IPOGK-----EAQYK-----ASTDFNSLFTTTNG---GRTY 101
DB 1049 EGEILLDGIQSGKYLKVKAPGEGVNISEYKEGKEITVSSGSEELLITIKAMKKGVI 1108
QY 102 VTCKDTASANI-----ATWAKISANTPVSSTVTSNN----- 135
DB 1109 LTKDSASHEVLADAEPELQNTAGSKLEKLTDDASGNIETDLAPGDYKLIETKAPTGY 1168
QY 136 --DGTEV-----INVSYQYVYVSVTVNGAVIM-----VTSVTPNATIIHKNTDATWG 182
DB 1169 QLDATPVNFTIDFNGQSEAAKVSKTNTAKTGTVMLTCKDSATNAELADATFELRNEDETLV 1228
QY 183 DGGKTVDPQKTVSYGCD-----TVKYTIYKNA-----VNVHGTPE 216
DB 1229 RENLVTDDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFITIDFNGQSEVSKVTNTA 1288
QY 217 KVVQYVI--KDTMPSASVVDL-----NEGS---YEVTITDGSNGNITTLTQSEKATGKN 266
DB 1289 KGTGTVLTCKDSATNAELADATFELRNESDILVRENLTDDNGEISV---ADLAPGDYK 1344
QY 267 LLENNNFTIIPWAATNPTGNTONGANDFFYKGINITITVYTVGLKSGAKPGSADLP 326
DB 1345 LIE-----TKAPAGYQLDAPVNF----- 1363
QY 327 ENTNIATINPNTSNDPPQKV-TVRDGOITIKKIDGSTKASIQAGIFVLKNGATGFLNPN 385
DB 1364 -----TIDFNGQSEAAKVSKTNTAKTGTVMLTCKDSATNAELADATFELRNESDILVREN 1417
QY 386 DTNNVWEGTEANATEYTTGADGIITGLKEGTYVYLVEKKAPLGYNL-----LDNSQ 437
DB 1418 -----LVTDNGEISVADLAPGDYKLIETKAPTGYQLDAPVNFIDFNQ 1462
QY 438 KVLGDGATDTTNSDNLVNV-PTVNNK-GTELPSTGGIGTTFIYIGAILVIGAGIV 493
DB 1463 SEAAKVTNKKKIGTIIVNFIDVNGQNDKEVHTGNVGE--YSVKAKEIVGHTLV 1518

RESULT 30
Q7WN54
ID Q7WN54 PRELIMINARY; PRT; 3346 AA.
AC Q7WN54;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Putative hemolysin.
GN BB1186.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdono-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagsis K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond B., Price C.,
RA Rabinowitsch E., Rutter S., Saunders M., Saunders D., Seeger K.,
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 10:26:31 ; Search time 19 Seconds  
(without alignments)  
1364.012 Million cell updates/sec

Title: US-10-009-254-2

Perfect score: 2578

Sequence: 1 MKKQMIQSLVSLAFGMV.....GAILVIGAGIVLARRRLRS 502

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 90 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/iaa/PCITUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	13.0	627	4	US-09-071-035-218
2	334	13.0	659	4	US-09-134-000C-6124
3	318	12.3	664	4	US-09-107-532A-7252
4	274.5	10.6	560	4	US-09-071-035-220
5	234.5	9.1	341	4	US-09-107-532A-5384
6	219	8.5	915	4	US-09-206-942-35
7	219	8.5	1222	4	US-09-206-942-37
8	219	8.5	1228	4	US-09-206-942-34
9	208.5	8.1	477	4	US-09-134-000C-6123
10	208.5	8.1	660	4	US-09-107-532A-6715
11	205.5	8.0	1027	4	US-09-107-532A-6675
12	199	7.7	508	4	US-09-107-532A-5331
13	194	7.5	1638	4	US-09-071-035-258
14	194	7.5	1638	4	US-09-071-035-258
15	194	7.5	1638	4	US-09-071-035-258
16	192.5	7.5	1095	4	US-09-206-942-43
17	192.5	7.5	1101	4	US-09-206-942-43
18	192	7.4	1220	4	US-09-206-942-26
19	192	7.4	1226	4	US-09-206-942-26
20	192	7.4	1747	4	US-09-134-000C-5999
21	190.5	7.4	1073	4	US-09-206-942-49
22	190.5	7.4	1079	4	US-09-206-942-47
23	190	7.4	669	4	US-09-071-035-264
24	186.5	7.2	1742	4	US-09-386-962C-4
25	185.5	7.2	1026	1	US-08-194-290-7
26	183.5	7.1	2736	4	US-09-252-991A-30227
27	183	7.1	2411	4	US-09-268-347-36

#### ALIGNMENTS

#### RESULT 1

US-09-071-035-218  
; Sequence 218, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:

28	183	7.1	3892	4	US-09-328-352-5503	Sequence 5503, Ap
29	181.5	7.0	1315	4	US-09-200-650E-5	Sequence 5, Appli
30	181.5	7.0	1833	4	US-08-621-944A-4	Sequence 4, Appli
31	181.5	7.0	1833	4	US-08-945-567D-4	Sequence 4, Appli
32	181.5	7.0	1992	4	US-08-621-944A-3	Sequence 3, Appli
33	181.5	7.0	1992	4	US-08-945-567D-3	Sequence 3, Appli
34	181.5	7.0	2048	4	US-08-268-347-48	Sequence 48, Appli
35	181	7.0	673	4	US-09-107-532A-5134	Sequence 5134, Ap
36	180.5	7.0	413	4	US-09-107-532A-6433	Sequence 6433, Ap
37	180.5	7.0	1180	4	US-09-206-942-65	Sequence 65, Appl
38	180.5	7.0	1188	4	US-09-206-942-63	Sequence 63, Appl
39	180	7.0	1166	4	US-09-200-650E-7	Sequence 7, Appli
40	180	7.0	1231	3	US-08-904-263A-4	Sequence 4, Appli
41	180	7.0	1231	4	US-09-434-123A-4	Sequence 4, Appli
42	180	7.0	2353	3	US-09-377-155-33	Sequence 33, Appl
43	180	7.0	2353	3	US-08-913-942-4	Sequence 4, Appli
44	180	7.0	2353	4	US-09-669-974-33	Sequence 33, Appl
45	180	7.0	2353	4	US-09-797-862-33	Sequence 33, Appl
46	180	7.0	2354	4	US-09-268-347-47	Sequence 47, Appl
47	179.5	7.0	905	4	US-09-134-000C-4420	Sequence 4420, Ap
48	177.5	6.9	1183	2	US-08-447-031A-2	Sequence 2, Appli
49	177	6.9	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
50	175.5	6.8	1536	1	US-08-038-682-2	Sequence 2, Appli
51	175.5	6.8	1536	1	US-08-302-832-2	Sequence 2, Appli
52	175.5	6.8	1536	2	US-08-530-138-2	Sequence 2, Appli
53	175.5	6.8	1536	2	US-08-469-880-2	Sequence 2, Appli
54	175.5	6.8	1536	2	US-08-728-470-2	Sequence 2, Appli
55	175.5	6.8	1536	2	US-08-617-697-2	Sequence 2, Appli
56	175.5	6.8	1536	3	US-08-719-641-2	Sequence 2, Appli
57	175	6.8	1095	4	US-09-206-942-69	Sequence 69, Appl
58	175	6.8	1536	4	US-09-206-942-67	Sequence 67, Appl
59	175	6.8	2314	4	US-09-268-347-49	Sequence 49, Appl
60	174	6.7	1026	3	US-08-614-377A-7	Sequence 7, Appli
61	174	6.7	1026	3	US-09-142-648B-7	Sequence 7, Appli
62	174	6.7	1233	4	US-09-134-000C-4971	Sequence 4971, Ap
63	174	6.7	1912	1	US-08-409-995-4	Sequence 4, Appli
64	174	6.7	1912	3	US-08-685-467-4	Sequence 4, Appli
65	173.5	6.7	952	4	US-09-107-532A-4706	Sequence 4706, Ap
66	173	6.7	1004	4	US-09-268-347-30	Sequence 30, Appl
67	173	6.7	1338	2	US-08-728-470-9	Sequence 9, Appli
68	173	6.7	1338	3	US-08-719-641-9	Sequence 9, Appli
69	173	6.7	1416	4	US-09-071-035-404	Sequence 404, App
70	173	6.7	1448	4	US-09-071-035-402	Sequence 402, App
71	173	6.7	1529	2	US-08-728-470-10	Sequence 10, Appl
72	173	6.7	1529	3	US-08-719-641-10	Sequence 10, Appl
73	173	6.7	1599	2	US-08-617-697-9	Sequence 9, Appli
74	173	6.7	1600	2	US-08-617-697-10	Sequence 10, Appl
75	172.5	6.7	1104	4	US-09-268-347-28	Sequence 28, Appl
76	172.5	6.7	1104	4	US-09-268-347-34	Sequence 34, Appl
77	172.5	6.7	2032	4	US-09-071-035-458	Sequence 458, App
78	172.5	6.7	2032	4	US-09-071-035-462	Sequence 462, App
79	172.5	6.7	2032	4	US-09-071-035-466	Sequence 466, App
80	172.5	6.7	2054	4	US-09-134-000C-6612	Sequence 6612, Ap
81	171	6.6	2142	4	US-09-540-236-3459	Sequence 3459, Ap
82	169	6.6	2504	4	US-09-328-352-5821	Sequence 5821, Ap
83	168	6.5	1132	4	US-09-198-452A-4961	Sequence 4961, Ap
84	167	6.5	507	4	US-09-134-000C-4961	Sequence 4961, Ap
85	166	6.4	1612	1	US-08-169-927-2	Sequence 2, Appli
86	164.5	6.4	558	4	US-09-071-035-268	Sequence 268, App
87	164	6.4	1098	1	US-08-409-995-2	Sequence 2, Appli
88	164	6.4	1098	3	US-08-685-467-2	Sequence 2, Appli
89	164	6.4	1098	3	US-09-377-155-32	Sequence 32, Appli
90	164	6.4	1098	3	US-08-913-942-2	Sequence 2, Appli



Db 627 STGGKGIYVVLGSGAVLLLIAGVYFARRK 656

RESULT 3

US-09-107-532A-7252

Sequence 7252, Application US/09107532A

Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Axiniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 7252:

SEQUENCE CHARACTERISTICS:

LENGTH: 664 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURES:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...664

SEQUENCE DESCRIPTION: SEQ ID NO: 7252:

US-09-107-532A-7252

Query Match 12.3%; Score 318; DB 4; Length 664;

Best Local Similarity 24.2%; Pred. No. 2.1e-16;

Matches 141; Conservative 91; Mismatches 206; Indels 144; Gaps 28;

QY 3 KKMIOQLLVASLAFMGMAVSPVTPIAFAA--ETGTTIV---QDTOKGATYKAYKVFDAEI 56

Db 137 KQAVQSL-----TGTTPVAGSTTDAGNVTLSPKKQNGKDAVY----- 175

QY 57 DNANVSDSNKGAS-----YLI PQKEAEYKAST---DFNSLF--TTTNGGRT 100

Db 176 ---TIKEEPKDGVSAAANNVLAPPVYEMIKQADGSYKGTBEIDTILYPKVTGNDGTL 232

QY 101 VYTKDGTASANEIATWAKSISANT--TP-----VSTVTB----- 132

Db 233 KVTIKIGTAE-NEALANGAEFIISKEEGTSPVKYIQSVTDGLXTWTITDQTKAKHFITGHSY 291

QY 133 --SNND-----GTEVINVSQYGYVYS--STVNGGAVIMVTSVTPNATHEKNT--- 177

Db 292 DTGNDDFAEASIEKGQLIVNHLEVGKYNLEEVKAPDNAMEIKQTIITPPEILANSQTPVE 351

QY 178 -----DATWGGGGKTVQDKTYTSVGDVTVKYITTYK---NAVNYHGTEKY-QYVIKOTMP 228

Db 352 KTIKNDTSKVDKTTPLQNGKDVAIKEKIQYEISVNIPLGIADKEGTQNKYTTFKLIDTHD 411

QY 229 SASVVDL-NEGSYEVITIDGSGNITLTQSGSEKATGKYNLLSENNFTITIP--WAAINT 285

Db 412 AALTFDNDSSGTAYALYDGNKEIDPV-----NYSVTQTDGFTVSDPNPIPSLT 462

QY 286 PTGNTQNGANDFFYKGINITITVYTVGLKSGAKPGSADLPENTNIATINPNTSNDDPGQ 345

Db 463 PGG-----TLKFVYVYMHLEKADPTKG-----FSNQANVDNGHTNDQTPP 502

QY 346 KTVRDGQITTKIDG--STKASLQGAIFVLKNA---TGQFLNFN-DTNNVEM-GTEANA 398

Db 503 SVDVVTTGGKRFVKVGDVTSQTLAGAEFVVVRDQSDSTAKYLSIDPSTKAVSWSAKESA 562

QY 399 TEYTTGADGIITITLKGTYVLEKAPLGNLLDNSQKVLGIDGATDTTNSDNLVNP 458

Db 563 TVFTTTSNGLIDVTGLKYGTYLBTETKAPEKVPVLTNRVAFITDEQSVYTAGQ---LISP 619

QY 459 --TVENNKGTSLPSTGGIGTTFIYIIGAILVIGAGIVLVARR 498

Db 620 EKIPNKHGKT-LPSTGGKGIYVVLGSGAVLLLIAGLYFARRK 660

RESULT 4

US-09-071-035-220

Sequence 220, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brooks

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 220:

SEQUENCE CHARACTERISTICS:

LENGTH: 560 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-220

Query Match 10.6%; Score 274.5; DB 4; Length 560;

Best Local Similarity 25.3%; Pred. No. 3.9e-13;

Matches 136; Conservative 72; Mismatches 212; Indels 117; Gaps 25;



Db 309 AKDNSNLITGDSNAGNTDAKKV-----TFSNVKDSKISASDHNVTLNSKVETSGDSTD 363  
QY 87 -----FNSLFTTTTNGRVTYTKDTPASAN-----EATWAKSISANTTP 126  
Db 364 EDGNNNTGLTITAKNVTVNNNITSHKTVNITASENVTTKAGTTINATTSVEVTAKTGD 423  
QY 127 VSTVTESSNDGTEV-----INVSQGYGYYVSSVNNGAVIMVTSVTPNATHEKNTDAT 180  
Db 424 IKGGIESNGVNIITASGDTLNVSNITQNVTVAAASGAVTTTKGSTINATNGNANITTK 483  
QY 181 WDGGGKTVQDKTYSVGDTVK-YTITYKNVNVHCTEKVYQVVIKDTWPSASVVDLNEGS 239  
Db 484 TGEINGE-VKSAGNVNITASGDTLNVSNITQNVTVTANSgai--TTTEGSTINATNGD 540  
QY 240 YEVITIDG-----SGNITTLTQSEKATGKYNLLENNNFTITIPMAATNPTGNT 290  
Db 541 ANITQTQNGINGKVESSESVTLIATGQTLAVGNIS-----GDTVITITADKGLTTQTSK 596  
QY 291 QNGANDDPFYKGI-----NTITVTVTVGL-----KSGAKPGSADLPENT 329  
Db 597 INGT-----KSVTSSQSGDISGTISGNTVSVSATGSLTTQAGSKIEAKTGEANVTSAT 650  
QY 330 NI--ATINPNTSNDPPQKVTVRDGOITIKKIDGSTKASLOGAIFVLKNATGQFLNFDNT 387  
Db 651 GTIGTISGNTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGLTTTETS 700  
QY 388 NNVEWG-----TEANATEYTTGADGIITITGLK-----EGTYVLVEKKA 426  
Db 701 SDITSSNGQTLTAKDSSISAGSINAANVTLNTTGT--LTTVAGSKIEAASGTLVINAKDA 758  
QY 427 PLGYNLLDNSOKVILGDGATDTNSDLLVNPVTVENNKGTELPSGTGGITTFIFYIGAIL 486  
Db 759 QL-----DGA---ASGDHTVNVNATNANGSGSVIATT-----SSRVNITGDLI 797  
QY 487 VIGAGIVLVAR 497  
Db 798 TIN-GLNIISK 807

RESULT 7  
US-09-206-942-37  
; Sequence 37, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 37  
; LENGTH: 1222  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-37

Query Match 8.5%; Score 219; DB 4; Length 1222;  
Best Local Similarity 22.5%; Pred. No. 2.5e-08;  
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;  
QY 29 AAGTGTITVQDTQKATYKAYKVFDAEIDNANVSNSKDGASYLIPQKKEAYKASTD-- 86  
Db 616 AKDNSNLITGDSNAGNTDAKKV-----TFSNVKDSKISASDHNVTLNSKVETSGDSTD 670  
QY 87 -----FNSLFTTTTNGRVTYTKDTPASAN-----EATWAKSISANTTP 126  
Db 671 EDGNNNTGLTITAKNVTVNNNITSHKTVNITASENVTTKAGTTINATTSVEVTAKTGD 730

QY 127 VSTVTESSNDGTEV-----INVSQGYGYYVSSVNNGAVIMVTSVTPNATHEKNTDAT 180  
Db 731 IKGGIESNGVNIITASGDTLNVSNITQNVTVAAASGAVTTTKGSTINATNGNANITTK 790  
QY 181 WDGGGKTVQDKTYSVGDTVK-YTITYKNVNVHCTEKVYQVVIKDTWPSASVVDLNEGS 239  
Db 791 TGEINGE-VKSAGNVNITASGDTLNVSNITQNVTVTANSgai--TTTEGSTINATNGD 847  
QY 240 YEVITIDG-----SGNITTLTQSEKATGKYNLLENNNFTITIPMAATNPTGNT 290  
Db 848 ANITQTQNGINGKVESSESVTLIATGQTLAVGNIS-----GDTVITITADKGLTTQTSK 903  
QY 291 QNGANDDPFYKGI-----NTITVTVTVGL-----KSGAKPGSADLPENT 329  
Db 904 INGT-----KSVTSSQSGDISGTISGNTVSVSATGSLTTQAGSKIEAKTGEANVTSAT 957  
QY 330 NI--ATINPNTSNDPPQKVTVRDGOITIKKIDGSTKASLOGAIFVLKNATGQFLNFDNT 387  
Db 958 GTIGTISGNTVN-----VTANTDNLTIK--DGARIKATGGAVTL--TATGGLTTTETS 1007  
QY 388 NNVEWG-----TEANATEYTTGADGIITITGLK-----EGTYVLVEKKA 426  
Db 1008 SDITSSNGQTLTAKDSSISAGSINAANVTLNTTGT--LTTVAGSKIEAASGTLVINAKDA 1065  
QY 427 PLGYNLLDNSOKVILGDGATDTNSDLLVNPVTVENNKGTELPSGTGGITTFIFYIGAIL 486  
Db 1066 QL-----DGA---ASGDHTVNVNATNANGSGSVIATT-----SSRVNITGDLI 1104  
QY 487 VIGAGIVLVAR 497  
Db 1105 TIN-GLNIISK 1114

RESULT 8  
US-09-206-942-34  
; Sequence 34, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 1228  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-34

Query Match 8.5%; Score 219; DB 4; Length 1228;  
Best Local Similarity 22.5%; Pred. No. 2.5e-08;  
Matches 124; Conservative 69; Mismatches 224; Indels 134; Gaps 23;  
QY 29 AAGTGTITVQDTQKATYKAYKVFDAEIDNANVSNSKDGASYLIPQKKEAYKASTD-- 86  
Db 622 AKDNSNLITGDSNAGNTDAKKV-----TFSNVKDSKISASDHNVTLNSKVETSGDSTD 676  
QY 87 -----FNSLFTTTTNGRVTYTKDTPASAN-----EATWAKSISANTTP 126  
Db 677 EDGNNNTGLTITAKNVTVNNNITSHKTVNITASENVTTKAGTTINATTSVEVTAKTGD 736  
QY 127 VSTVTESSNDGTEV-----INVSQGYGYYVSSVNNGAVIMVTSVTPNATHEKNTDAT 180  
Db 737 IKGGIESNGVNIITASGDTLNVSNITQNVTVAAASGAVTTTKGSTINATNGNANITTK 796

```

; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...660
; SEQUENCE DESCRIPTION: SEQ ID NO: 6715:
US-09-107-532A-6715

Query Match      8.1%; Score 208.5; DB 4; Length 660;
Best Local Similarity 23.4%; Pred. No. 6.5e-08;
Matches 127; Conservative 62; Mismatches 190; Indels 163; Gaps 25;

QY 10 LVASLAF-----GNAVSPVPTPIAFAAETGITTVDDTQGATYKAYKVFDAAEIDNANV 61
    |||:|||::|||::|
Db 210 IVSKLSFLTKFGDGDGTGAQSKQTPL-----KGAKFELYK-----243

QY 62 SDSNKDGASYIIPOCKEAERYKASTDFNSLETTTNGRTYVT-----KKDTASA 110
    :||:|||||:|
Db 244 ---GEPG-----KGTGLDLIVSDQQKLTAIDLTLGKYTFVEVPSEVVVGSKDEPTAQ 294
    :||:|||||:|

QY 111 NEIATWAKSISANTTPVSTVTSNNDG--TEVINVSQYGYY--YVSSFTVNNGAVIMVTSVT 167
    :||:|||||:|
Db 295 YLLGADARNDAHN----KLAFEITNDGTSDLKASVYNYPAPDKTWTNG-----341
    :||:|||||:|

```



```
QY 168 PNATHEKNTDWTGDDGGKTVQDKTYSYVGDVTKYTIITYNNAVNYHTEKYVYVYKDTM 227
Db 342 -----TGOEHSFQIGD-----AVNYQGT-----IHIPTDIA 367
QY 228 PSASVVDLNEGSVEVT-----ITDGSNITTLTQSEKATGKYN--LLBENNFTIT 277
Db 368 CGADGITVNGVKSETSPYSVFKWGTAGQGLSVAAKANIKVTKDGSVVLKENTDYKIQ 427
QY 278 IPWAATNTPTGNTONGANDPFYKGINITVTYTVGVKSGAK-PSGADLPENTNIATINP 336
Db 428 -----NSENGVIDFI---VNNQVSDTVASLHGQDLQMTYNNVYVNDSSAAVNP 473
QY 337 NTS-----NDP-----CQKTVRDQGITIKID-----GSTKASLOGAIFVLKN 376
Db 474 LTNVDVFNPNFQOEHEKTKADVYTGAKFLKVDGSLGFTGIGATPESABFAKN 533
QY 377 ATGQF---LNFNDTNVVE---WGTE-ANATEYTTGADGIIITITGLKEGYIYLVKAPLG 429
Db 534 ASGKYGGLVDTDKNGKVAVVDVANAAILKSKDKEGHFEITGLTEGEYSLEETKAPEN 593
QY 430 YNLLNSQKVLGDGATDNTSDNLLVNPVTNNKGTLPSTGGIGTTFYIIGAILVIG 489
Db 594 YQKL---TKEISFKVDKDSYKEENRI---TIKNQKASVPMTGSNGFQYVLISCLL-LG 646
QY 490 AG 491
Db 647 AG 648

RESULT 11
US-09-107-532A-6675
; Sequence 6675, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; FILING DATE: 30-Jun-1998
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 6675:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1027 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
```

```
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1027
; SEQUENCE DESCRIPTION: SEQ ID NO: 6675:
US-09-107-532A-6675

Query Match      8.0%; Score 205.5; DB 4; Length 1027;
Best Local Similarity 21.7%; Pred. No. 2.1e-07;
Matches 127; Conservative 67; Mismatches 179; Indels 211; Gaps 26;

QY 23 VTPIAFAAETGTTIVQ--DTOKGATY---KAYKVPDAEIDNANVSDSNKDGASYLIPQSK 77
Db 424 VDKVSFNKETNLVVDVFGNLAEGQSIFYEYAEVTDVDFN---EGROEDGFS----- 473
QY 78 EAEYKASTDFNSLFTTTTNGGRTYVTK-----KDTASANE-IATWAKSI 120
Db 474 ---RNEASMTYNNQVMS-----RDYELRIPESNLPALSRKSVDKVINLENELIYTLHV 526
QY 121 S-----ANTPVSVTBESNNDGTEVINVSQGYVYVSVTSVNGAVIMVTSVTPNATIHK 175
Db 527 SMFSGTRNTLVIEDVLE---EGMEFVSFLSDIPSIDAEVNG---IVRFSANNTI--- 576
QY 176 NTDATMGDGGGKT-----VDQKTSYSGDVTYKTYITIKYNAVNYHTEKYVYVYKDTMP 228
Db 577 -----GGPTSFTLSFKVNVEAYVGDQILNFAFLVNSDNRLNTSTVTRKI----- 622
QY 229 SASVVDLNEGSVEVTITDGSNITTLTQSEKATGKYNLLE-----ENNPF-T 275
Db 623 -----DGRKVIKTD-----ENGQRLAGAKFEILSSDSQVIOEGITAEENGFLS 666
QY 276 ITIPWAA-----TNPPTGNTONGANDPFYK--GINTITVTYTVGVKSGA-----KPGS 322
Db 667 EPLPVGYSQVRETEAPEGYLLDATVHDVFINESYIEEVTLTITENILDTGGVELIKHAEGS 726
QY 323 ADL-----PENTNIATINPN-TSNDP-----PG----- 344
Db 727 GEVLQAGVFNQNRREGETLQGLTTGDEGKLAIDGLAPGAYQLVETQAPIGYELDATPIE 786
QY 345 -----QKVTVRDQGITIKIDGSTKASLOGAIFVLKNATGQFLNFDNTNNV 390
Db 787 FEIERSQTAVVVELTKENRLTPGGVVLTKIDQSGEILQAGVFNQNRREGETLQ----- 839
QY 391 EWGTEANATEYTTGADGIIITITGLKEGYIYLVKAPLGYNL-----LDNSQKVLIG 442
Db 840 -----TGLTTGDEGKLAIDGLAPGTQVQLVETQAPIGYELDATPIEPIERSQTAV-- 889
QY 443 DGATDTTNSDNLVNPVTNNKGTLPSTGGIGTTFYIIGAIL 486
Db 890 -----VELTKENRLTPGGVVLTKIDQSGEIL 916

RESULT 12
US-09-107-532A-5331
; Sequence 5331, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GFC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 5331:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...508
; SEQUENCE DESCRIPTION: SEQ ID NO: 5331:
US-09-107-532A-5331

Query Match
Best Local Similarity 7.7%; Score 199; DB 4; Length 508;
Matches 118; Conservative 84; Mismatches 216; Indels 130; Gaps 23;

QY 1 MKKKMQSLVASLAEQMAVSPVTPAFATETGTVQ-----DIO 41
DB 36 MKKLGLWSMCLLFLFPAPFQV-----ATETEMVQIHLKLLPFGNOLPKRPNNDGQ 90
QY 42 KGATYKAYKVFDAEIDNANVSDNKGASYLIPQKGEAEYKASTDFNSLPTTTTNGGRTV 101
DB 91 EKALLQYR-----GLNGVTFQV-----YDVTDSEVHL-----REKKTIV 125
QY 102 VTKDPTASANEIATWAKSISANTPTVSTVTESNNDGTEVINVS-----QYGYVVSST 154
DB 126 ----EEAQA-EIAKNGASSGMFTAETTTLLNNDGSIASFSLAAKQEKRDKAYLFIESK 180
QY 155 V-----NNGAVIVTSVTPN---ATIH--EKNTDATWGDGGKTV---DQKTVSGD 198
DB 181 VPEVVEKAENMVVLPHVHQQNKKUSTIHLXPNEENDYPDPPEKVLPEEPRNDFTE 240
QY 199 TVKYTTITYKNVNYHGTEKYQYVVKIDTMPASVVDLNEGSYEVVITDGSIGNITTLTQGS 258
DB 241 KITVSLHTTIPVNIIDYQK---FELSDSDEA-----LTFLP---NSLTSSNG 283
QY 259 EKATGKNLLENNFTI--TIWPAATNPTGNTQNGANDDFPYKGINITVITYTGVLKS 316
DB 284 EKLTEGVIIHKHGHGFDVLSFPLEK-----YAG-KKLTISYQMLSS 326
QY 317 GAKPGSADLPENTNIATINPTNSD---DPGOKVTVRQGIITIKID-GSTKASIQGAI 371
DB 327 TAQANK-----EINNGILDFGFGVSTKKVSVYTGSKQFVKLETKNPKDKRLAGAV 376
QY 372 FVLKNATGQFLNFDNTNNVW-GTEANATBYTTGADGIITIGLKEGTYYLVEKKAPLGY 430
DB 377 FLIKRAGNYLQ-QTANGYKWTKNESDALHLISDKNGAFSISGLKTSYRLKEIAPSGY 435
QY 431 NLLDNSQKVLGQADTTNSDLLNVPVNNKGTLEPSTGGIGTITFIIGAILVIGA 490
DB 436 ILSETEIPFTISTFLSEDKADSLKVNKKENSRLPFLPKTNETKNTLLGVGVGMVFASFA 495
QY 491 GIVLVARR 498
DB 496 IWLFIKKR 503
;
; RESULT 13
; US-09-071-035-258
; Sequence 258, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 258:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-258

Query Match
Best Local Similarity 7.5%; Score 194; DB 4; Length 1638;
Matches 136; Conservative 56; Mismatches 219; Indels 212; Gaps 26;

QY 23 VTPIAFAETGTITVQDTQKGATYKA---YKVFDAEIDNANVSDNKGASYLIPQKGEA 79
DB 340 VSNIFYAAGTTSEVTAPTITVTGNSTAGYEVKGTADANATVEIRNAGGT--VIGTG--- 394
QY 80 EYKASTDFNSLPTTTTNGGRTVTKDPTASANEIATWAKSISANT-----TPVSVTV 131
DB 395 ----TADGTGAFVTVPAGE-----AGANETLTAVAKNASGTEXTPTTFTQTPADEAT 442
QY 132 ESNNDGTEVINVSQYGVYVVSSTVNGAVIMVTSVTPNATHEKNTDAT-----WGDG-G 185
DB 443 VTAPTITVTGNSTAG-YEVKGTAD-----ANATVEIRNAGGTVIGTGADGTG 490
QY 186 GKTVDQKTVSGDVTVKYTTITYKNVNYHGTEKYVQ-----YV 222
DB 491 AFTVTVPAGEAGANETLTAVAKNASGTESTPTTFTQTPADEATVAPTITVTGNSTAGYE 550
QY 223 IKDTPSASVVDLNE-----GSVEVITDG-SGNITTLTGSEKATGKNYL 267
DB 551 VKGTADANATVEIRNAGGAVIGTGADGTGAFVTVPIPAGEAGANETLTAVAKNASG--- 606
QY 268 LEENNNFTITIWAATNTP-----TGNTQNGANDDFYKGINITITVT----- 309
DB 607 -TESTPTTFTQTP-ADPNTFPVATPIVETVTGTTKG-----YEVKGTAEVGTITIEVRDAAGT 660

```

QY 310 -----YGVLSKSG-----AKGSDLPENTNIATNPNT 338  
DB 661 VLGTATTGDKYTVTLDSGTATANTQTLVAVAKNASGTSQATATTPADVTAPVDNIT 720  
QY 339 SNDDPGQKVT-VRDQGIITKKIDGSKASLOGAIF--VLKNATGOF-----LNFNDDT 387  
DB 721 GNSGSGYEITGADPNNTTIEVRDPS-----GAVIGTGTSDANGDFTVTLPTGTNPNGDT 774  
QY 388 NNVT-----EWGTEANATEYTTGADGIIT-----ITGLKEGTYYLVEKKAPLGYNLLDNSQ 437  
DB 775 LTVIGKDNAGNESQPTVELVLPADATVTAFTVGTGNSVAGYQVTGTADPNATIEIRDAD 834  
QY 438 KVLGDGATDTTNS-----ITGLKEGTYYLVEKKAPLGYNLLDNSQ 460  
DB 835 GNVIATGTADGTGSGFAVNLPAGTANANETLTALAKDPAGNTSTPTTFTQPADEVVAPPSV 894  
QY 461 E-----NNKGTLPSTGGIGTTI 478  
DB 895 DKVTGNTTQGYQVGTGAELGTTI 917  
  
RESULT 14  
US-09-071-035-262  
; Sequence 262, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 262:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1638 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-262  
  
Query Match 7.5%; Score 194; DB 4; Length 1638;  
Best Local Similarity 21.8%; Pred. No. 3.3e-06;  
Matches 136; Conservative 56; Mismatches 219; Indels 212; Gaps 26;  
  
QY 23 VTFPIAFAETGRTITVQDTQKGYAKA---YKVFDAEIDNANVSDSNKDGASYLIPOGKEA 79  
DB 340 VSNIFYAAGTSEVTAPITITGVTGNSVAGYQVGTADANATVEIRNAGT--VITG--- 394  
QY 80 EYKASTDFNSLFTTTTNGGRTVTVTKKDTASANEIATWAKSISANT-----TPVSTVT 131

DB 335 ----TADGTGFTVTPAGE-----AGANETLTAVAKNASGTEXTPTTFTQPADBAT 442  
QY 132 ESNNDGTEVINVSQYGYVSVSNVNGAVIMVTSVTPNATIEHKNTDAT-----WGDG-G 185  
DB 443 VTAPITITGVTGNSVAGYQVGTADANATVEIRNAGT--VITG--- 490  
QY 186 KTVUDQKYSVGDVTKYITTYKNAVNYHGTKEVYO-----YV 222  
DB 491 AFTVTPPAGEAGANETLTAVAKNASGTESTPTTFTQPADATVTAFTVGTGNSVAGYQVGTADANATVEIRNAGT--VITG--- 550  
QY 223 IKDTMPSASVVDLNE-----GSEVETITDG-SGNIITLTQSGSEKATGKYNL 267  
DB 551 VAGTADANATVEIRNAGGAVIGTGTADGTGFTVTPPAGEAGANETLTAVAKNASG----- 606  
QY 268 LEENNNFTITTPWAATNP-----TGNTONGANDDFYKGINITITVT----- 309  
DB 607 -TESTPTTFTQTP-ADPNTPVATPIVETVGTGTTKG-----YEVKGTAEVGTITIEVRDAAGT 660  
QY 310 -----YGVLSKSG-----AKGSDLPENTNIATNPNT 338  
DB 661 VLGTATTGDKYTVTLDSGTATANTQTLVAVAKNASGTSQATATTPADVTAPVDNIT 720  
QY 339 SNDDPGQKVT-VRDQGIITKKIDGSKASLOGAIF--VLKNATGOF-----LNFNDDT 387  
DB 721 GNSGSGYEITGADPNNTTIEVRDPS-----GAVIGTGTSDANGDFTVTLPTGTNPNGDT 774  
QY 388 NNVT-----EWGTEANATEYTTGADGIIT-----ITGLKEGTYYLVEKKAPLGYNLLDNSQ 437  
DB 775 LTVIGKDNAGNESQPTVELVLPADATVTAFTVGTGNSVAGYQVTGTADPNATIEIRDAD 834  
QY 438 KVLGDGATDTTNS-----ITGLKEGTYYLVEKKAPLGYNLLDNSQ 460  
DB 835 GNVIATGTADGTGSGFAVNLPAGTANANETLTALAKDPAGNTSTPTTFTQPADEVVAPPSV 894  
QY 461 E-----NNKGTLPSTGGIGTTI 478  
DB 895 DKVTGNTTQGYQVGTGAELGTTI 917  
  
RESULT 15  
US-09-071-035-266  
; Sequence 266, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:  
; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/071,035  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: A. Anders Brookes  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB369P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 262:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1638 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-071-035-262

```
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 266:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-266

Query Match          7.5%; Score 194; DB 4; Length 1638;
Best Local Similarity 21.8%; Pred. No. 3.3e-06;
Matches 136; Conservative 56; Mismatches 219; Indels 212; Gaps 26;

QY 23 VTPFAFAAETGTTTVDQTKGATYKA---YKVPDAEIDNANVSDSNKDGASYLIPQGEKA 79
Db 340 VSNIFYAAGTTSVTAPTTGTGNSAGVEYKGTADANATVEIRNAGT--VIGTG--- 394
QY 80 EYKASTDFNSLFTTTTNGGRYVTKDTSANEIATWAKSISANT-----TPVSTVT 131
Db 395 ----TADGTGAFVTVTPAGE-----AGANETLTAVAKNASGTEXTPTTQTPADEAT 442
QY 132 ESNNDGTEVINVSQYGVYVSVTVNNGAVIMVTSVTPNATIEHKNTDAT-----WGDG-G 185
Db 443 VTAPTITGVGNSTAG-YEVKGTAD-----ANATVEIRNAGGTGVTGTADGTG 490
QY 186 GKTVDQKTSVGDTVKYTITYKNAVNYHGTKEVYQ-----YV 222
Db 491 AFTVTVPAGEAGANETLTAVAKNASGTESTPTTFTQTPADEATVTAPTITGVGNSTAGYE 550
QY 223 IKDTPSASVVDLNE-----GSYEVTTIDG-SGNITTLTGSEKATKYNL 267
Db 551 VKGTADANATVEIRNAGGAVIGTGTADGTGAFVTVTPAGEAGANETLTAVAKNASG---- 606
QY 268 LEENNNFTIIPWAAFNTP-----TGNTONGANDDFYKGINITVT----- 309
Db 607 -TESTPTTQTP-ADPNPVPATPIVETVGSITKG-----YEVKGTAEVGTITVEVRAAGT 660
QY 310 -----YTVLKGSG-----AKPGSADLPENTNIATINPT 338
Db 661 VLGTATTTGTDGKYVTLLDSGTATANTQSLVAVAKNASGTESQAPATATTPADVTAFTVDNIT 720
QY 339 SNDDPGQKVT-VRDQGITIKKIDGSPKASLOGAIF--VLKNATQF-----LNFNT 387
Db 721 GNSGSGYEITGTADPNNTIEVRDPS-----GAVIGTGTSDANGDFTVLPTGTTPNGDT 774
QY 388 NNV-----EMCTEANATEYTTGADGIIT-----ITGLKEGTYYLVVEKKAPLGYNLLDNGS 437
Db 775 LTVIGKDNAGNESQPTIEVLVPADATVTAFTVGTGNSVAGYQVGTADPNATIEIRDAD 834
QY 438 KVLGDPGATDTNS-----DNLLVNPVT 460
Db 835 GNVIATGTDGTSFAVNLPAGTANANETLTALAKDPAGNTSTPTTFTQTPADEVWAPPSV 894
QY 461 E-----NNKTELPSTGGIGTTI 478
Db 895 DKVTGNTTQGYQVGTGAELGTTI 917

RESULT 16
US-09-206-942-45
; Sequence 45, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loomsore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; EARLIER FILING DATE: 1998-12-08
```

```
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 1095
; TYPE: PRP
; ORGANISM: Haemophilus influenzae
US-09-206-942-45

Query Match          7.5%; Score 192.5; DB 4; Length 1095;
Best Local Similarity 23.8%; Pred. No. 2.4e-06;
Matches 128; Conservative 63; Mismatches 194; Indels 153; Gaps 30;

QY 40 TOKGATYKAYKVFDAEIDNA-----NVSDNKGASLYLP---QGKEAEYKASTD-FNSLIF 91
Db 511 TNSDSAYRT--IIEGNTNSNGDLNITD-NKNNAEIOIGGNISQKEGNLTITSSDKINITN 567
QY 92 TTTTNGGRYVTKDTSANEIATWAKSISANTTPVSTVTVESNNDGTEVINVSQYGVYV 151
Db 568 QITIKKG---VNKEDSDST-----ANNANLTITKTELQLTGD---LNIS--GFDKA 611
QY 152 SSTVANGAVIM-----VTSVTPNATIEHK-----NTDA-- 179
Db 612 EITAKEGADLLIGSDNNDNNANAKVTFNOVKDSKISADSHNVTLNSKVETSGNNDAES 671
QY 180 TWGDGGKTVDOKTVSGDVKYTYKNAVNYHGTKEVYQVVIKDTWPSASVVDLNEGS 239
Db 672 NNGDGTSLTINAKNI---TVNNNITSHKTVNITASENV-----TTKAGTTINATGGS 720
QY 240 YEVTTIDG-----SGNITTLTGSEKATKYNLLENNNFTIIPWAAFNTPGT 290
Db 721 VEVTAKTGDIKKVESTSGSVTLTATGEALAVSNIS---GNTVTITANKGKLTQAGST 776
QY 291 QNGANDDFYKGI-----NTIITVYTVGL--KSG-----AKPGSADLPENT 329
Db 777 VSAIN-----GVTASSQSGDISGTSIGNTVKVSAGIDLTTKSGSEIRAKTGEANVTSAT 830
QY 330 NI--ATINPNTSNDPQKQVTVRDQGITK---KIDG-----STKASLOGA 370
Db 831 GTIGTISGNVAVN-----VTANTGLTIVEDAAKIDATGGAATLTATSGKLTAKSSS-- 882
QY 371 IFVLKNATGQFLNFDNNVEMG--TEANATEYTTGADGIITIG---LKEGTYYLVVEK 424
Db 883 ----ITSANNO-VNLSAKDGSIGGNINAAVTLNTTGA--LITVKGSSINANSGLTIVINAK 936
QY 425 KAPLGYNLLDNSQKV-----ILGDGATDTNSDNLVNPVTVENNKGTELPSTGGIGTTI 478
Db 937 DAELNGEASGNHTVVNATNANGSGSVIATTSRNVITGDLITINGLNIISKNGINTVL 994

RESULT 17
US-09-206-942-43
; Sequence 43, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loomsore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 1101
; TYPE: PRP
; ORGANISM: Haemophilus influenzae
US-09-206-942-43
```

Query Match 7.5%; Score 192.5; DB 4; Length 1101;  
Best Local Similarity 23.8%; Pred. No. 2.4e-06;  
Matches 128; Conservative 63; Mismatches 194; Indels 153; Gaps 30;

QY 40 TQKATYKAYKVFDAEIDNA-----NVSDNKGASYLIP---QKGEAEYKASTD-FNSLIF 91  
DB 517 TNSDSAYRT--IIEGNIINSNGDLNITD-NKNNASIQGGNISQKEGNLTSSDKINITN 573  
QY 92 TTTTNGGRTYVTKDITASANEIATWAKSISANTTPVSTVTSNNDGTEVINVSQGYVVV 151  
DB 574 QITIKKG--VNKEDSDST-----ANNANLTIKTELQLTGD-----LNIS--GFDKA 617  
QY 152 SSTVANGAVIM-----VTSVTPNATHEK-----NTDA-- 179  
DB 618 EITAKEGADLIIGNSDNNNNNAKVFNQVKDSKISADSHVTLNSKETSNGNDAES 677  
QY 180 TWGDDGGKTVDOKTVSGVTYKYNVNHGTEKVVYQVVIKDTMPSASVVDLNEGS 239  
DB 678 NNGDGTSLTINAKNI-----TVNNNITSHKTVNITASENV-----TTKAGTTINATGGS 726  
QY 240 YEVVITDG-----SGNIITLTOGSEKATGKYNLLENNNFTITIPWAATNPTGNT 290  
DB 727 VEVTAKTGDIKGKVESTGSVTLTATGEALAVSNIS-----GNTVITANKGKLJTQAGST 782  
QY 291 QNGANDDPFKGI-----NTIIVTYTGVL--KSG-----AKPGSADLPENT 329  
DB 783 VSAIN-----GVTASSQSGDISGTISGNTVKSVAIGDLTTKSGSEIKAKTGEANVTSAT 836  
QY 330 NI--ATINPNTSNDPQKQVTVRDGQITIK---KIDG-----STKASLOQA 370  
DB 837 GTIGGTISGNAV-----VTANTGDLTVEDDAKIDATGGAATLTATSGKLTTKASSS-- 888  
QY 371 IFVLKNATGQFLNFNDNNVNG--TEANATEYTTGADGIITG-----LKEGTYYLVVEK 424  
DB 889 ---ITSANNQ-VNLSAKDGISGGINAANVTLLTTGA--LTTVKGSSINANSGLTLVINAK 942  
QY 425 KAPLGYNLLDNSQV-----ILGDGATDTNSDNLVNPVENNKGTELPSTGGIGTTI 478  
DB 943 DAELNGEASGNHTVNVNATNANGSGSVIATTSRNVITGDLITINGLNIISKNGINTVL 1000

RESULT 18  
US-09-206-942-28  
; Sequence 28, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 1220  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-28

Query Match 7.4%; Score 192; DB 4; Length 1220;  
Best Local Similarity 23.2%; Pred. No. 3e-06;  
Matches 119; Conservative 61; Mismatches 203; Indels 130; Gaps 23;

QY 55 EIDNANVSDNKGASYLIPQKGEAEYKASTDFNSLFTTTTNGGRTYVTKDITASANEIA 114  
DB 660 ETSNGNDAESNNGDS-----TSLTINAKNVTVNNNITSHKTVNITASEN-VT 706  
QY 115 TWAKS-----ISANTTPVSTVTSNNDGTEV-----INVSOGYTYVVSSTVNG 158

DB 707 TKAGTTINATIGSVEVTAKTGDIKGGIESNGNVTITASGDTLNVSNITGQNTVVAASG 766  
QY 159 AVIMVTSVTPNATHEKNTDATWGDGGKTKVDQKTVSVGDTVK-YTITYKNAVNVHGTGTEK 217  
DB 767 AVTTTKGSTINATTGNANITTKTGEINGE-VKSASGNVNITASGNTLNVSNITGQNTVTV 825  
QY 218 VYQVVIKDTMPSASVVDLNEGSYEVVITDG-----SGNIITLTOGSEKATGKYNLL 268  
DB 826 ANSGAI--TTTEGSTINATTGDANITTTQGTNGINGKVESSSGVTLLIATGQTLAVGNIS-- 881  
QY 269 EENNFITIPWAATNPTGNTQNGANDDPFKGI-----NTIIVTYTGV 313  
DB 882 --GDTVITADKGLTKTQTSKINGT-----KSVTSSQSGDISGTISGNTVSVATGS 933  
QY 314 L-----KSGAPGSADLPENTNI--ATINPNTSNDPQKQVTVRDGQITIKIDGSKTA 365  
DB 934 LTTQAGSKIEAKTGEANVTSATGTIGGTISGNTVN-----VTANTDNLTIK--DGARIK 985  
QY 366 SLOGAIFVLKNATGQFLNFNDNNVNG-----TEANATEYTTGADGI 408  
DB 986 ATGAVTL--TATGGTLLTTETSSDITSSGQTTILTAKDSSIAGSINAAVNTLNTTGT--L 1041  
QY 409 ITITGLK---EGTYVLVEKKAPLGYNLLDNSQKVLGDGATDTNSDNLVNPVENNK 464  
DB 1042 TTVAGSKIEAASGLTLVINAKDAQL-----DGA---ALGDRTEVNTNANGS 1084  
QY 465 GTELPSTGGIGTTIYIIGAILVIGAGIVLVAR 497  
DB 1085 GSVIATT---SSRVNITGDLITIN-GLNIISK 1112

RESULT 19  
US-09-206-942-26  
; Sequence 26, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; EARLIER FILING DATE: 1998-12-08  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 26  
; LENGTH: 1226  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-26

Query Match 7.4%; Score 192; DB 4; Length 1226;  
Best Local Similarity 23.2%; Pred. No. 3.1e-06;  
Matches 119; Conservative 61; Mismatches 203; Indels 130; Gaps 23;

QY 55 EIDNANVSDNKGASYLIPQKGEAEYKASTDFNSLFTTTTNGGRTYVTKDITASANEIA 114  
DB 666 ETSNGNDAESNNGDS-----TSLTINAKNVTVNNNITSHKTVNITASEN-VT 712  
QY 115 TWAKS-----ISANTTPVSTVTSNNDGTEV-----INVSOGYTYVVSSTVNG 158  
DB 713 TKAGTTINATIGSVEVTAKTGDIKGGIESNGNVTITASGDTLNVSNITGQNTVVAASG 772  
QY 159 AVIMVTSVTPNATHEKNTDATWGDGGKTKVDQKTVSVGDTVK-YTITYKNAVNVHGTGTEK 217  
DB 773 AVTTTKGSTINATTGNANITTKTGEINGE-VKSASGNVNITASGNTLNVSNITGQNTVTV 831  
QY 218 VYQVVIKDTMPSASVVDLNEGSYEVVITDG-----SGNIITLTOGSEKATGKYNLL 268

Db 832 ANSGAI--TTTEGSTINATTGDANITTTQTNNGKVESSGSVTLIATGQTLAVGNIS-- 887  
Qy 269 EENNFTITIPWAAATNPTGNTONGANDFFYKGI-----NTITVTYTG 313  
Db 888 --GDTVIITADKGLITQTSKINGT-----KSVTSSOSGDISGTISGNTVSVSATGS 939  
Qy 314 L-----KSGAKPGSADLPENTNI--ATINPNTSNDPQGVTRVDCQITIKKIDGSTKA 365  
Db 940 LTTQAGSKIEAKTGEANVTSATGTIGGTISGNTVN-----VTANTDNLTIK--DGARIK 991  
Qy 366 SLOGAIFVLKNATGQFLNFDNTNNVWG-----TEANATEVTTCGADGI 408  
Db 992 ATCGAVTL--TATGGTLTTTSSDITSSNGQTLTAKDSSIAGSIANAANVTLLNTTGT--L 1047  
Qy 409 ITITGLK---EGTYLVEKAPGLGYNLLDNSQKVLGDGATTTNSDNLNLTNNPTVNNK 464  
Db 1048 TVVAGSKIEAASGTLVINAKDAQI-----DGA---ALGDRTEVNVNTWANGS 1090  
Qy 465 GTPLPSTGGIGTTFYIIGAILVIGAGIVLVAR 497  
Db 1091 GSVIATT---SSRVNITGLIITN-GLNIISK 1118

RESULT 20  
US-09-134-000C-5999  
; Sequence 5999, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134, 000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055, 778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5999  
; LENGTH: 1747  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5999

Query Match 7.4%; Score 192; DB 4; Length 1747;  
Best Local Similarity 21.7%; Pred. No. 5.1e-06;  
Matches 136; Conservative 59; Mismatches 234; Indels 198; Gaps 26;

Qy 11 VASLAFGMVSPVTPIAFAETGTTIVQDTQKGATYKA---YKVFDAEIDNANVSDSKD 67  
Db 439 VAKNASGTESTPTT-FQTPADEATVTA-PTITGVTGNSTAGYEVKGTADANATVEIRNAG 496  
Qy 68 GASYLIPQKGEAEYKASTDPNSLFTTTTNGGRVTVTKKOTASANEIATWAKSISANTTPV 127  
Db 497 GT--VIGTG-----TADGTGAVTVVPAGEAGANETLTAVAKNASGTESTPTTQTPA 547  
Qy 128 STVTESNNDGTEVINISQYGYVSVSVVNGAVIMVTSVTPNATIEHKNTDAT-----WG 182  
Db 548 DEATVTAPTITGVTGNSTAG-YEVKGTAD-----ANATVEIRNAGGTVIGTGA 595  
Qy 183 DG-GGKTVQDKTSVSGTVKYTTYKNAVNYHGTCKYQ-----220  
Db 596 DGTGAFVTVVPAGEAGANETLTAVAKNASGTESTPTTFTQTPADEATVTAPTITGVTGNET 655  
Qy 221 --VVIKDTMPSASVDLNE-----GSYEVTITDG-SGNITTLTGSEKATG 263  
Db 656 AGEVKGTDANATVEIRNAGGAVIGTADGTGAVTVIPAGEAGANETLTAVAKNASG 715  
Qy 264 KYNLLBNNTTITIPWAAATNP-----TGNTONGANDFFYKGINTIIVT-----309  
Db 716 ----TESTPTTQTP-ADNTEPVATPIVETVIGSTTKG---YEVKGTAEVGTITIEVRD 765  
Qy 310 -----YTGVLKSG-----AKPGSADLPENTNIATI 334

Db 766 AAGTVLGTATTGCDGKYIVTLLDSTGATANTQTLSSVAKNASGTESQPATATTADVTAPTV 825  
Qy 335 NPNNTSNDPQKQVT-VROGQITIKKIDGSTKASLOGAIF--VLKNATGQF-----LN 383  
Db 826 DNITGNSGSGYEITGADPNTTIEVRDPS-----GAVIGTGTSDANGDFTVLTPTGTTN 879  
Qy 384 FNDTNNV---BWSGTEANATEVTTGADGIIT-----ITGLKEGTYYLVEKAPGLYNLL 433  
Db 880 PGDTLTVIIGKNAGNESQFTEVLVPADATVTAPTVTGVTGNSVAGYQVTGADPNATIEI 939  
Qy 434 DNSQKVLGDGATDTTNS-----DNLIV 456  
Db 940 RDADGNVIATGTADCTGSGFAVNLPAGTANANETLTALAKDPAGNTSTPTTQTPADEVVA 999  
Qy 457 NPTVE-----NNKGTELPSTGGIGTTI 478  
Db 1000 PPSVDKVTGNTTQGYQVGTGTAELGTTI 1026

RESULT 21  
US-09-206-942-49  
; Sequence 49, Application US/09206942  
; Patent No. 6432669  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H.  
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High  
; FILE OF INVENTION: Molecular Weight Proteins  
; FILE REFERENCE: 1038-861 MIS:jb  
; CURRENT APPLICATION NUMBER: US/09/206,942  
; CURRENT FILING DATE: 1998-12-08  
; EARLIER APPLICATION NUMBER: 09/167,568  
; EARLIER FILING DATE: 1998-10-07  
; NUMBER OF SEQ ID NOS: 95  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 49  
; LENGTH: 1073  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-206-942-49

Query Match 7.4%; Score 190.5; DB 4; Length 1073;  
Best Local Similarity 21.1%; Pred. No. 3.3e-06;  
Matches 126; Conservative 71; Mismatches 197; Indels 203; Gaps 28;

Qy 9 LLVASLAFGMVSPVTPIAFAE-----TGTTT-----VDTQKQKATYKAYKVFDAEI 56  
Db 447 LVGASANTNNLSVKSGAKFKAETNDNLNITGTTTNGTSTIIDVKKGAA-----KLGNTN 502  
Qy 57 D-NANVSDSKGASLYLPQKGEAEYKASTDPNSLFTTTTNGGRT-----YVTKKD--- 106  
Db 503 DGNLNTTNNAKQKQSVI--NGNITNNKALNI-----TNNNDTEIOIGNISOKENL 555  
Qy 107 TASANEI--ATWAKSISANTTPVSTVTESNNDG-----TEVINVSQYQ 148  
Db 556 TISSDKNITKRIEIKAGT-----DQNSDSGVSANLANLTIKKELKTENLNIISGFDK 609  
Qy 149 YVSVSTVNGAVI-----MVTSTVTPNATIEHK-----NTDA 179  
Db 610 AEIVAKENNLIIGNNGNDNANAKTVPFNNVKDSKISANGHNVTNLNSKVETSDGNSNTEG 669  
Qy 180 TWGDCGGKTVQDKTSVSGDTVKYTTYKNAVNYHGTCKYQYVVIKDTMPSASV-----232  
Db 670 NSDNNAGLTIDAKV-----TVNNDITSHKTVNITASERI-----DTKADTTINATGNV 719  
Qy 233 -----VDLNEGSYEVTITDG-----SGNIT-----TLTGSEKATGKY 265  
Db 720 KLTAVTSDIOGGIKNSGDVNITTTSTGSKGSKSVTLTATEKTLTVGN-----772  
Qy 266 NLENNNTTITIPWAAATNP-----TGNTONGANDFFYKGI-----NTITVTY 310

Db 773 ---VSGNTVTVTANRGALTTLAGSTINGTN-----GVTTSSQSGEIGGEVTKTVSVTA 823

Qy 311 TG---VLKSGAKPGSADLPENTNIATINPNTSNDPQGVKQVTVRQGIITIKKIDGSTKASL 367

Db 824 TAGSLTVKGGAK---INATEGTATLTASSCKLTVTEASSNITSAGKQVDLSAQDGS-----875

Qy 368 QGAIFVLKNATQGFNFNDTNVWGTENATEYTTGADGIITIG-----LKEGTYYLVE 423

Db 876 -----IAGQI-----SAANVTLLNTTGT--LTTVEGSSINANEGLTVINA 912

Qy 424 KKAPLGYNLLDNSQV-----ILGDGATDTTNSDNLNVPNTVENNKGTLPSTGGIGT 476

Db 913 NDAKLDGKASGRNTEVNATNAGSGSVTAKTSSSVNITGDLNTINGLNIISENGRT 969

RESULT 22

US-09-206-942-47

; Sequence 47, Application US/09206942

; Patent No. 6432669

; GENERAL INFORMATION:

; APPLICANT: Loomore, Sheena M.

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H.

; TITLE OF INVENTION: Protective Recombinant Haemophilus influenzae High

; TITLE OF INVENTION: Molecular Weight Proteins

; FILE REFERENCE: 1038-861 MIS:Jb

; CURRENT APPLICATION NUMBER: US/09/206,942

; CURRENT FILING DATE: 1998-12-08

; EARLIER APPLICATION NUMBER: 09/167,568

; EARLIER FILING DATE: 1998-10-07

; NUMBER OF SEQ ID NOS: 95

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 47

; LENGTH: 1079

; TYPE: PRT

; ORGANISM: Haemophilus influenzae

US-09-206-942-47

Query Match 7.4%; Score 190.5; DB 4; Length 1079;

Best Local Similarity 21.1%; Pred. No. 3.3e-06;

Matches 126; Conservative 71; Mismatches 197; Indels 203; Gaps 28;

Qy 9 LLVASLAFGMVSPVPIAFAB-----TGIT-----VDTQKATYKAYKVFDAEI 56

Db 453 LVGASANINNNLSVKSGAKFAETDNLNITGTFTNNGTSIIDVKKGA-----KLGNITN 508

Qy 57 D-NANVSDNKGASVLIPOGKEAEYKASTDFNSLFTTTTNGRT-----VYTKD--- 106

Db 509 DGNLNTTNNAKQKSVI-NGNITNNKALNI-----TNNGNDTEIQIGGNISQKGNL 561

Qy 107 TASANEI-ATWAKSISANTTPVSTVTESNNDG-----DOGNSDGVASNANLTIKTKELKLTENLISGDPK 515

Db 562 TISSDKINIKRIEIKAGT-----DOGNSDGVASNANLTIKTKELKLTENLISGDPK 515

Qy 149 YVVSSTVNNGAVI-----MVTSVTPNATHEK-----NTDA 179

Db 616 ABIVAKENNNLIIGNNGDNANAKVTFNNVVKDSKISANGHNVLNLSKVTSDGNSNTEG 675

Qy 180 TWGDDGGKTVDOKTSYVSGVTYKTYLYKAVNVHGTKEYVQYVVKDTPMSASV-----232

Db 676 NSDNNAGLTIDAKV-----TVNNDTSHKTVNITASERI-----DTKADTINATNGV 725

Qy 233 -----VDLMESGYEVTITDG-----SGNIT-----TLTQGSERATGY 265

Db 726 KLTAVTSDIQGGKNSGDVNITTSNGSINGKIESKSGSVTLTATEKTLTVGN-----778

Qy 266 NLLSENNFTIIPHAATVPTGNTONGANDOFFYKGI-----NITVTV 310

Db 779 ---VSGNTVTVTANRGALTTLAGSTINGTN-----GVTTSSQSGEIGGEVTKTVSVTA 829

Qy 311 TG---VLKSGAKPGSADLPENTNIATINPNTSNDPQGVKQVTVRQGIITIKKIDGSTKASL 367

Db 830 TAGSLTVKGGAK---INATEGTATLTASSCKLTVTEASSNITSAGKQVDLSAQDGS-----881

Qy 368 QGAIFVLKNATQGFNFNDTNVWGTENATEYTTGADGIITIG-----LKEGTYYLVE 423

Db 882 -----IAGQI-----SAANVTLLNTTGT--LTTVEGSSINANEGLTVINA 918

Qy 424 KKAPLGYNLLDNSQV-----ILGDGATDTTNSDNLNVPNTVENNKGTLPSTGGIGT 476

Db 919 NDAKLDGKASGRNTEVNATNAGSGSVTAKTSSSVNITGDLNTINGLNIISENGRT 975

RESULT 23

US-09-071-035-264

; Sequence 264, Application US/09071035

; Patent No. 6448043

; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi

; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION NUMBER: US/09/071.035

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: A. Anders Brookes

; REGISTRATION NUMBER: 36,373

; REFERENCE/DOCKET NUMBER: PB369P2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

; TELEFAX: (301) 309-8512

; INFORMATION FOR SEQ ID NO: 264:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 669 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-071-035-264

Query Match 7.4%; Score 190; DB 4; Length 669;

Best Local Similarity 22.5%; Pred. No. 1.8e-06;

Matches 122; Conservative 60; Mismatches 242; Indels 118; Gaps 22;

Qy 11 VASLAFGMVSPVPIAFABETGTTVDTQKATYKA---YKVFDAIDNANVSDSKD 67

Db 13 VAKASGTEXTPTT-FQTPADENVTA-PTTGTGTCSTAGYEVKGTADANATVEIRNAG 70

Qy 68 GASVLIPOGKEAEYKASTDFNSLFTTTTNGRTYVTKDTSANEIATWAKSISANTTPV 127

Db 71 GT--VIGTG-----TAGTGAFVTVPAGEAGANETLTAVAKNASGTSTPTTFFQTPA 121

Qy 128 STVTESNNDGTEVINVSQGYVYVSSTVNGGAVIMVTSVTNATHEKNTATWGDGSGK 187

Db 122 DEATVAPTITGVTGNSTAG-YEVKGTADANATVEIRN-AGGAVIGTGTADGT-----GAF 175

Qy 188 TVDQKTSYVSGTVKTYITYKNAVNVHGTKEYVQ-----YVI 223

Db 176 TVTIPAGEAGANETLTAVAKNASGTSTPTTFFQTPADNPVATPVTGTTGTTKGYEV 235

[illegible]

		7.2%; Score 185.5; DB 1;	Length 1026;
		Best Local Similarity 22.1%; Pred. No. 7.6e-06;	
		Matches 131; Conservative 66; Mismatches 240;	Indels 157; Gaps 26
Qy	6 IQSLVLASLAFGMVSPVTPIAFAETGTTVDOTKGATYKAYKVDAEINANVNVSQN	65	
Dd	299 VQAAVTALPTGVTSIGIE--TMNVTSGAAITLTNTSSGT-----GLTALTNTN	344	:::
Qy	66 KDGA SYLI PQGEAE YKASTDFNSLF TTTNGG-----RTYYTK-KDTASANEIATWAK	118	:
Dd	345 TSQAQT VTAGAGNL TATTAA QAANNVA VDGGANVT VASTGTVS GTTCTTVGANSA ASGTV	404	: :
Qy	119 SIS -ANTTPVST-----VTESNNDCTEVINVSQGY	148	: :   : :



Db 405 SVSVANSSTTTTGAIVTGGTAVTAQTAGNAVNTTLTQADVTVTGNSSTTAVTQTQAA 464  
QY 149 YVVSSTVN---NGAVIMVTSVTPNATIHEKNTDATTWGDGGKTVQDKTVSGDVTVKYIT 205  
Db 465 ATAGATVAGRVNGAVTITDSAAASATTAGKIAATVTLGSGCAATIDSSALT----- 514  
QY 206 YKNAVNHGTEK---VYQVVIKDTMPSASVVDLNEGSYEVT---ITDGGG---NITTLTQ 256  
Db 515 ---TVNLSGTGSLGIRGALTAI-PTANTLTILNVNGLTTGAITDSEAAADGGFTIMI 570  
QY 257 GSEKATGKYNLENNNFTITIPMAATNTPTGNTQNGANDDFYKGINITVT----- 309  
Db 571 AGSTASSTIASVAADATTILNSGDARVITSHTA-----AALTGIVTNSVGATL 621  
QY 310 -----YTGVLKSGAKPOSADLPENTNIA-----TINPNTSNDPQGVKTVRVDG-QIT 356  
Db 622 GAELATGLVFTGAGAGADSILLGATTKAIVMGAGDDTVTVTSATLGGGSGVNGDGDVLV 681  
QY 357 KKIDGSTKASLOGAIFVLKNATQCF-----LNFNNTNNVNGTEANATEYTT 403  
Db 682 ANVNGSS-----PSADPAFGFETLTVAGAAAGGSHNANGFTALQIGATAGATTFTN 733  
QY 404 GADGI-ITITGLKEGYIYVEKAPLG-----YNL-LDNSQKVLGD-----GAT 446  
Db 734 VAVNGLTVLAAPTGTITVTLANA-TGTSDFVNLTLSSAALAAGTVALAGVETVNIAT 792  
QY 447 DTNNS---DNLLVNPT-----VENNKGTELPSTGGIGTTFIYIGAILVIGAG 491  
Db 793 DTNNTAHVDTLTLQATSAKSIYVTVGNAGLNLNTGNTAVTSF---DASAVTGTG 843

## RESULT 26

US-09-252-991A-30227  
; Sequence 30227, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30227  
; LENGTH: 2736  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-30227

Query Match 7.1%; Score 183.5; DB 4; Length 2736;  
Best Local Similarity 20.7%; Pred. No. 4.5e-05;  
Matches 112; Conservative 73; Mismatches 222; Indels 133; Gaps 21;  
QY 21 SPVTPIAFAAETGTI---TVQDTQKATYKAYKVFDA-----EIDN-ANVSDSNKDG 68  
Db 1066 TPSTPLA---DGTVNATATDPAGNTGGGTTVDIAIPATPTVNLSSGSLSGTAEPG 1121  
QY 69 ASYLIPQKAEYKASTDFNSLFTTTNGGR---TYVTKKDTASANEIATWAKSISANTTP 126  
Db 1122 STVILTGD-----NGNPFAEVATDGGSNWYTPSTPIANGTVVNVVAAADAGNSSP 1172  
QY 127 VSTVTT-ESNNDGTEVINVSQYGYVVSSTVNVGAVIMVT-----SVTPNATHEKNT 177  
Db 1173 PATVTVDSSAPPAPVNPIS--NGWISGTAEGATVTLTDAGNPNIGQVATDAGSGNWSF 1230  
QY 178 DATWGDGGKTVQDKTVSGDVTVKYITITVKNANVNHGTEKVVQYVIKDTMPSASVVDLNE 237  
Db 1231 PGTPPANGTIVATATDPTGNTGPQAAVTVDAV-----APPAPVIDPSN 1274

QY 238 GS-----YEVITDGSNIITLTQSEKATCKYNLLENNNFTITIPHAATNPTPG 288  
Db 1275 GTTISGTAEAGAKVILTDGNG--PIGETTADSGNMTFTPATPLANGTVVNAVAQDPAG 1332  
QY 289 NT--QNGANDDFYKGINITITVTVGLKSGAKPGS----- 322  
Db 1333 NTGPGSITVDVAVPNTPVNPSNGNLLNGTAEPGSTVLTLDGNGNPIGQTTADSGNWS 1392  
QY 323 ---ADLPENTNI-ATINPNTSNDPQGVKTVRVDGQITIKKIDGSTKASLOGAIFVLKNA 377  
Db 1393 FTPGSQLPNGTVVNVTAASDAAGNTSAPATTVTDSSLPSPQVDPNSGVSIGTA----- 1446  
QY 378 TGQFLNPNFNTNVNVEWGTANAT---EYTTGADGIITIT---GLKEGTYLVEKKAPLGN 431  
Db 1447 -----DAGNTIITIDGNGNPIGQVATDAGSGNWSFTPGIPLDGTVVNVVARSF---S 1495  
QY 432 LLDNSOKVILGDGATDTTNSDNLLNVPVNNKGTETPSTGGIGTTFIYIGAILVIGAG 491  
Db 1496 NVDSAPAVITVDGVAAPAA-----PVIDPSNGTEISGTAEGATV-----ILTDGGG 1541

## RESULT 27

US-09-268-347-36  
; Sequence 36, Application US/09268347  
; Patent No. 6335182  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M.  
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
; FILE REFERENCE: 1038-860  
; CURRENT APPLICATION NUMBER: US/09/268,347  
; CURRENT FILING DATE: 1999-03-16  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 36  
; LENGTH: 2411  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-09-268-347-36

Query Match 7.1%; Score 183; DB 4; Length 2411;  
Best Local Similarity 22.1%; Pred. No. 4.1e-05;

Matches 124; Conservative 71; Mismatches 194; Indels 172; Gaps 30;

QY 32 TGTITVQDTQKATYKAYKVFDA--EIDNANVSDSNKDGASYLIPQKAEYKASTDFNS 89  
Db 190 SSSFTPNDEKTRAATVKDVLNAGWNIKGAKTAGGNVESVD-LVSAYNNVEF-ITGDXT 247  
QY 90 L---FTTTTNGGRITV--TKKDTA---SANEIATWAKSISANTTPVSTVTSNNNDGTEVI 141  
Db 248 LDVVLTAKEGKTEVKTFTPKTSVIKEDGKLFTGKENNDTNKVTSTNTATDNTDEGNLV 307  
QY 142 N-----VSQVGYVVSSTVN--NGAVIMVTSVTPNATIEHKNTDATWGDGGKTVDQK 192  
Db 308 TAKAVIDAVNKAHVRKTTTANGQGDPAVTSGT-VNTPES-----GDGTTASVTKD 359  
QY 193 TYSVGDVTKYTITVKNANVNHGTEKVVQYVIKDTMPSASVVDL-----NEGSEYV 242  
Db 360 TNGNGITVKYDAKVGDLGKLFDSDKKIVADTTALTITVGKVAEIAEKDDKKLVNAGDLVT 419  
QY 243 TI-----TDGSGNITTLTGSE-----KATGKYNLLENNNFTITIPWA 281  
Db 420 ALGNLSWKAKEADTTDGALEGISKDQEVKAGETVTFKAGKNLKVQDGANFTYSLQDA 479  
QY 282 ATNPT---GNTONGAND-----DFFYKGINITITVTVGLKSGAKPGS 322  
Db 480 LTGLTSITLGGTTGNDKATVINKDGLTITIPAGNGTTGNTISVT-----KDGIKAGN 534  
QY 323 ADLPENTNIAT-----INPNTSNDP-----GQ 345  
Db 535 KAI---TVVASGLRAYDDANFDVLNNSATLNRHVEDAYKGLLNLNEKNANKQPLVTDST 591  
QY 346 KTVTRDQGITTKKIDGSTKASLOGAIFVLKNATQGFNFDNNTNVEWGTANATEYTTGA 405

Db 592 AATVGD----LRKL-----GWVSTKNGTKE-----ESNQVKQADEV-----LFTGA 629  
QY 406 DGIITITGLKEG-----TYLVEKKAPLGYNLLDNSOKVILGDGAT-----DTTNSDNL 455  
Db 630 -GAATVTSKSENGKHTITVSVAETKADSG-----LEKGDITIKLVQNDQNDVL 678  
QY 456 VNPTVENKGTETLPSTGGIGT 476  
Db 679 ---TVGNN-GTAV-TKGGFET 694

RESULT 28  
US-09-328-352-5503  
; Sequence 5503, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; CURRENT APPLICATION NUMBER: GTC99-03PA  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 5503  
; LENGTH: 3892  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-5503

Query Match 7.1%; Score 183; DB 4; Length 3892;  
Best Local Similarity 23.8%; Pred. No. 8.2e-05;  
Matches 149; Conservative 69; Mismatches 218; Indels 190; Gaps 35;

QY 12 ASLAFGMVSPVPIAFAAE-TGFTVQDTQKGATYKAYKVEDAEI---DNANVSPNKD 67  
Db 240 ANVATVINGVYNATVDAAGTWTVSPEGSLVADADKTIIDAKVTFTDAAGNSSVND 299  
QY 68 GASYLE----PQGEAEYKASTDFNSLFTTTNGRTYVTKDTSANEIA----TWAKS 119  
Db 300 TQTVTLDTAPNAPVIDPVNGTD--PITGTAEPGSTVTVTPDGSKTIVVAGPDGTW--- 354  
QY 120 ISANTPVSVTESNDGTEVINVSQYGYVSVSTVNGGAVIMVTSVTPNATHEKNT-D 178  
Db 355 ----TVP-----NPLNDGDDEVATAT-----DPAGNTSGPATVADVAVPTVALDVLTD 402  
QY 179 ATWGDGGKTVDOKTSVSGTVKYTITYKNAVNHGTEKYVQVIXD-TMPSASVVDLNE 237  
Db 403 STPALTG--TVNDPTATVVNVND-GVDYPAVNGDGT-----WTLADNLTPT-----LAD 449  
QY 238 GSYEVITI--TDGSGNI-----TTLTQSEKATGKYNLLLENNNFTIT 277  
Db 450 GPHITITVATDAAGNVGTDGVTVDTAAPNTAGVTFTIDSVTADNVINASEAGNVIT 509  
QY 278 -----IPWAATNTP-----TGNTONGANDDPFKGINTITVTVGLKSGAK----- 319  
Db 510 GVLKNIPADATNATVTVINGVYNATVDK--TAGTWTVSVPGLVADADKTIIDAKVTF 567  
QY 320 ---PGSADLPENTNIATIT-----NPNSTNDPDQKVTV----- 349  
Db 568 TDAAGNSTVNDTQIYTLDTAAPAVVIDPVNGTDPITGTAEPGSTVTVTPNGDPTATVV 627  
QY 350 -----RDG-QITIKKIDGSKASLQGAIFVLKNATG---QPLNFNDTNVNE 391  
Db 628 AGPDGWSVNPGLNDGDEVEIATDPAGNPSLPGTATV--DAVGPNTDGVNFT-VDSVT 684  
QY 392 WGTENATYTTGADGIIITGLKEG-----TY-YLVEKKAPL----- 428  
Db 685 ADNVINASE-----ASGNVTITGLKNVPADAANTVTVWINGQTYTATVDSGTAGTVSV 740  
QY 429 -GYNLLDNSOKVI-----LGDGAT---DTTNSNLLVNPVTVENKGTETLP 469  
Db 741 PGSDLTADADKTIIDAKVTFTDAAGNSSVNDTQTYITIDTAPAPVINPV-----NGTD-P 795

QY 470 STG----GIGTTFIYIIGAILVIGAG 491  
Db 796 ITGTAEPGSTVTVTPDGSSTTTTVAG 821

RESULT 29  
US-09-200-650E-5  
; Sequence 5, Application US/09200650E  
; Patent No. 6680195  
; GENERAL INFORMATION:  
; APPLICANT: Patti, Joseph M.  
; APPLICANT: Foster, Timothy J.  
; APPLICANT: Hook, Magnus A.O.  
; APPLICANT: Bidhinn, Deirdre Ni  
; APPLICANT: Perkins, Samuel L.  
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus  
; FILE REFERENCE: P06283US2/BAS  
; CURRENT APPLICATION NUMBER: US/09/200,650E  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,815  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: 60/098,427  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1315  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-200-650E-5

Query Match 7.0%; Score 181.5; DB 4; Length 1315;  
Best Local Similarity 22.0%; Pred. No. 2.2e-05;  
Matches 116; Conservative 72; Mismatches 191; Indels 149; Gaps 29;

QY 34 TTVQDTQKGATYKAYKV--FDAETDNANVSDNSDKGAS---YLIPOGKEAB--YKAST- 85  
Db 427 TTVNPSSENSLTNAKLKVQAHSSYPN-NIGQINKDVTDIKIYQVPGYTLNKGVDVNTK 485  
QY 86 -----DFNSLFTTTTNGRTVV-----TKDTSANEIATWA--- 117  
Db 486 ELTDTVNTQYLKLTIGDNNNSAVIDFGNADSAVVWVNTKFQYTNSESTLVOATILSSG 545  
QY 118 -KSISANTTPSVTSTESNNDGTEVINVSQYGYVSVSTVNGGAVIM---VTSVTPNATI 172  
Db 546 NKSVSSTGNALGFTNNQSGAGQEVYKI---GNYVWEDTNKNGVQELGEGKGVNVT--VTV 600  
QY 173 HERNTDATWGDGGKTVDOKTSV-----GDTVKTIITYKN----- 208  
Db 601 FDNNTNTKVGE--AVTKEDGSYLIPNLPGD---YRVEFSNLPKGYEVTTPSKQGNNEEL 655  
QY 209 -----AVNYHGTEK-----VYQ-----YVIKDTMPSASVVDLNE---GSYEVTID 246  
Db 656 SNGLSASVITVNGKDNLSADLGIYKPKYNLGDVWEDTNKN-GIQODEKIGISGVTVTLKD 714  
QY 247 GSGNITLTLTQSEKATGKYNLLE-ENNNFTITIPWAATNTPGTNGTQNGANDDPFYKGINT 305  
Db 715 ENGNV--LKTVTITDADGKYKFTDLONGYKVEFTTPEGVTPITVT--SGSDIEKDSNGLT 771  
QY 306 IIVTVTVGLKSKAKPGSADLPENTNIATINPNTSNDPDQKV---TVRDQITIKKIDGS 362  
Db 772 -----TGVI-NGA-----DNMTLDSGFYKTPKYNLGNVWEDTNKDKG-----QDS 811  
QY 363 TKASLQGAIFVLKNATGQPLNFNDTNVWEGTEANATEYTTGADGIIITITGLKEGTYLIV 422  
Db 812 TEKIGISGVTVTLNENGEVLQTTKDK-----DKYQFTGLENGT-YKV 854  
QY 423 EKAPLGYNLLDNSOKVILGDGATDTTNSDNLVNPVTVENKGTETLP 470  
Db 855 EPETPSGYT-----PTQVGSCTDGEIDSGNSTTGTGVIKDKDNDTIDS 896

RESULT 30

US-08-621-944A-4  
; Sequence 4, Application US/08621944A  
; Patent No. 6440425  
; GENERAL INFORMATION:  
; APPLICANT: SASAKI, Ken  
; APPLICANT: HARKNESS, Robin E.  
; APPLICANT: LOOMORE, Sheena M.  
; APPLICANT: KLEIN, Michel H.  
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER  
; TITLE OF INVENTION: MEMBRANE PROTEIN OF MORAXELLA  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/621,944A  
; FILING DATE: 26-MAR-1996  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/478,370  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-587  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1833 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-621-944A-4

Query Match 7.0%; Score 181.5; DB 4; Length 1833;  
Best Local Similarity 23.7%; Pred.No.3.6e-05;  
Matches 136; Conservative 59; Mismatches 205; Indels 173; Gaps 31;  
QY 17 GMAVSPVTPFAFAAETGTTTQDQTKGATYKAYKVPDAEIDNANVSDSNKDGASYLIPQG 76  
Db 351 GISVTP-TEISVDKSGNVT-----APTYNI-----GVKTELNSDGT----- 388  
QY 77 KEAEYKASTDFNSLFT-----TTNGRGTYVTK-KDTASANEIATWAKSISA 122  
Db 389 DKFSVKSGTNSLSVTAELASLYLNEVNRNRTADSLQSFVKEEDDDANAI-TVAKDITK 447  
QY 123 NITPUSV-----TESNNDGTEVINVSQ-YGVYVVSSTVNGAV-----IMV- 163  
Db 448 NAGAVSILKLGKNGLTVAKKDGVTFGLSQDSGLTIGKSTLNNDGLTVKDTNEQIQVG 507  
QY 164 -----TSVTPNATIEHKNTDATWGGGKT-----VDQKTYSGVDTVKYTI 204  
Db 508 ANGKFTVNGSNPGTGIANTARIITDKIGFAGSDGAVDTNKPFLDQDKLVGN-VKITN 566  
QY 205 TYQNAVNYHGTKEYQYVTKDTPMSASVVDLNEGSYEV--TITD-----GSGNITTLTQGS 258  
Db 567 TGINA---GGRAITG--LSPTLP--SIADQSSRIELGNTIQDKKSNAAASINDILN-- 616  
QY 259 EKATGKYNLLENNNTIIPW-----AATNPTGNTONGANDFFYKGINITVT 309  
Db 617 ---TG-FNLKNNNPIDFVSTYDVIDFANGNATTATVTHDTANKTSKVVDYVNVDDTTIH 672

QY 310 YTGIV---LKSGAKPGSAD-LPENTNIATINPTNSDDPGQKVTYRD----- 351  
Db 673 LTGTTDDNKKLGKVTTKLNKTSANGNTAT-NFNVNSDEEDALVNAKDIAENLNTLAKEIHT 731  
QY 352 -----GQITIKIDGSKASLQGAIFV-LKNATGQF--LNFNDTNNVWEGTEANAT 399  
Db 732 TKGTTADTALQTFVTKVDENNADANAITVGQKANNQVNTLTILKGENGLNIKTDKNGT 791  
QY 400 EYTTGADGIIITGLKEGTYYLVEKKAPLGYNLLDNSQKVLGSGATDTTNSDNLVNPT 459  
Db 792 V-----TFGINTTSLKAG-----KSTLNDGGL-----S 815  
QY 460 VENNKGTELPSTGIGCTTIFYIIGAILVIGAGI 492  
Db 816 IKNPTGSEQIQVGADGVK-FAKVNNGVVGAGI 847

Search completed: July 22, 2004, 10:29:49  
Job time : 24 secs

